ALMA MATER STUDIORUM - UNIVERSITA DI BOLOGNA
In Cotutela Con:
KATHOLIEKE UNIVERSITEIT LEUVEN

DOTTORATO DI RICERCA IN
INGEGNERIA ELETTRONICA, INFORMATICA E DELLE TELECOMUNICAZIONI

CICLO: XXVIII

Settore Concorsuale Di Afferenza: 09/E3 Elettronica
Settore Scientifico Disciplinare: ING-INF/01 Elettronica

Developing Ultrasound-Based Computer-Aided Diagnostic Systems Through Statistical Pattern Recognition

Presentata da: **Mahdi Tabassian**

Coordinatore Dottorato:
Prof. Dr. **Alessandro Vanelli Coralli**

Relatore:
Prof. Dr. **Jan D'hooge**

Relatore:
Dr. Ing. **Luca De Marchi**

Correlatore:
Prof. Dr. **Guido Masetti**

ESAME FINALE ANNO 2016

To my beloved parents, brother and sister.

Acknowledgements

When I started to seriously think about applying for a Ph.D. position abroad, I knew that it won't be only an opportunity to achieve valuable scientific knowledge but also a great chance to being familiar with other cultures and people. Now, in the end of my Ph.D. journey, I believe I was lucky because I could experience working in two universities in two different countries, made wonderful friends and learned a lot of worthwhile things from them and their cultures.

First of all, I would like to sincerely thank my supervisors Prof. Jan D'hooge, Dr. Luca De Marchi and Prof. Guido Masetti. Jan, thanks a lot for letting me be a member of your excellent research team. Working with you was a great honour for me and I have benefited so much from your guidance, skills and the time that you have devoted for managing my research projects in Leuven. Luca, I would like to extend a special thanks to you for all your support and guidance. You were not only my supervisor but also a good friend who helped me a lot when I needed it.

I am grateful for the careful readings of the manuscript and the constructive comments by the external reviewers Prof. Olivier Bernard and Dr. Nicolas Duchateau and would like to thank the jury members Prof. Piero Tortoli and Prof. Matthew Blaschko.

I would like to acknowledge the opportunity provided for me by the university of Bologna to spend my internship at KU Leuven. I especially thank Prof. Alessandro Vanelli-Coralli, Ph.D. board coordinator, for his administrative support and cooperation. I also want to

thank staff and administrators of both the university of Bologna and KU Leuven.

To my colleagues and friends in Bologna. Alessandro, Giovanni, Nicola, Francesca, Ali, Mohsen, Erfan, Mohammadsadegh, Mohammad and Kourosh, thank you for sharing your time with me and for all your help. To my colleagues and friends in Leuven. Martino, thanks a lot for helping me being familiar with a new research line and enviroment, for your time, useful comments and also for being a great friend. I also would like to thank Carolina, Hang, Ling, Alejandra, Natalia, Bidisha, Adriyana, Margot, Oana, Natasa, Brecht, Raja, João, Vangjush, Nuno, Pedro Santos, Stathis, Jurgen, Frederic, Jeroen, Peter, Pedro Morais, Sandro, Ali, Ehsan, Esmaeel and Ahmadreza for the wonderful times together and for making my stay in Leuven memorable.

My especial thanks go to my best friend Mohammadreza for his unconditional support and for the unforgettable times during our trips.

Finally, I would like to express my deepest love and sincere gratitude to my family whose support, patience and encouragement during pursuit of my Ph.D. made this accomplishment possible.

Abstract

Computer-aided diagnosis (CAD) is the use of a computer software to help physicians having a better interpretation of medical images. CAD systems can be viewed as pattern recognition algorithms that identify suspicious signs on a medical image and complement physicians' judgments, by reducing inter-/intra-observer variability and subjectivity.

The proposed CAD systems in this thesis have been designed based on the statistical approach to pattern recognition as the most successfully used technique in practice. The main focus of this thesis has been on designing (new) feature extraction and classification algorithms for ultrasound-based CAD purposes. Ultrasound imaging has a broad range of usage in medical applications because it is a safe device which does not use harmful ionizing radiations, it provides clinicians with real-time images, it is portable and relatively cheap.

The thesis was concerned with developing new ultrasound-based systems for the diagnosis of prostate cancer (PCa) and myocardial infarction (MI) where these issues have been addressed in two separate parts. In the first part, 1) a new CAD system was designed for prostate cancer biopsy by focusing on handling uncertainties in labels of the ground truth data, 2) the appropriateness of the independent component analysis (ICA) method for learning features from radiofrequency (RF) signals, backscattered from prostate tissues, was examined and, 3) a new ensemble scheme for learning ICA dictionaries from RF signals, backscattered from a tissue mimicking phantom, was proposed. In the second part, 1) principal component analysis (PCA) was used

for the statistical modeling of the temporal deformation patterns of the left ventricle (LV) to detect abnormalities in its regional function, 2) a spatio-temporal representation of LV function based on PCA parameters was proposed to detect MI and, 3) a local-to-global statistical shape model based on PCA was presented to detect MI.

Contents

List of Figures

List of Tables

List of Acronyms

A ..... atrial filling
ACC ..... accuracy
ANN ..... artificial neural network
BBA ..... basic belief assignment
C ..... circumferential
CAB ..... computer-aided biopsy
CAD ..... computer-aided diagnosis
CLT ..... central limit theorem
CMR ..... cardiovascular magnetic resonance
CZ ..... central zone
D ..... diastasis
DRE ..... digital rectal examination
D-S ..... Dempster-Shafer
DWKNN ..... distance-weighted k-nearest neighbor
DWPT ..... discrete wavelet packet transform
E ..... early filling
ECG ..... electrocardiography
ED ..... end-diastole
Ejec ..... ejection
EMC ..... electromechanical coupling
ES ..... end-systole
FIS ..... neuro-fuzzy inference systems
FLD ..... Fisher linear discriminant
FN ..... false negative
FP ..... false positive

GDA ..... generalized discriminant analysis
GGD ..... generalized Gaussian distribution
IB ..... integrated backscatter
IC ..... independent component
ICA ..... independent component analysis
ISM ..... intercept, slope and midband fit
IVC ..... isovolumetric contraction
IVR ..... isovolumetric relaxation
IVUS ..... intravascular ultrasound
KNN ..... k-nearest neighbor
L ..... longitudinal
LA ..... left atrium
LMB ..... local mean-based
LV ..... left ventricle
MLP ..... multilayer perceptron
MRI ..... magnetic resonance imaging
PC ..... principal component
PCa ..... prostate cancer
PCA ..... principal component analysis
PDF ..... probability distribution function
PDM ..... point distribution model
PSA ..... prostate specific antigen
PZ ..... peripheral zone
R ..... radial
RA ..... right atrium
RF ..... radiofrequency
ROI ..... region of interest
RSM ..... random subspace method
rtCAB ..... real-time computer-aided biopsy
RV ..... right ventricle
SE ..... sensitivity
SNR ..... signal to noise ratio
SP ..... specificity
S-T ..... spatio-temporal
STI ..... speckle tracking imaging
SVM ..... support vector machines

TDI ..... tissue Doppler imaging
TN ..... true negative
TP ..... true positive
TRUS ..... trans-rectal ultrasound
TZ ..... transition zone
ULA-OP ..... ultrasound advanced open platform

Chapter 1

Introduction

1.1 Computer Aided Diagnosis

The use of a computer software to help physicians having a better interpretation of medical images is called *computer-aided diagnosis* (CAD). More specifically, CAD algorithms analyze medical images to give an estimate of the likelihood that an image represents a specific disease process [14, 33]. These CAD algorithms, however, should be considered as tools that complement physicians' judgments, by reducing inter-/intra-observer variability and subjectivity, and not as substitutions of specialists' experience.

CAD systems can be viewed as pattern recognition algorithms that identify suspicious signs on a medical image and bring them to the attention of the physician with the aim of reducing the false negative readings [14]. In this thesis, the proposed CAD systems have been designed based on the statistical approach to pattern recognition as the most intensively studied and successfully used technique in practice [56].

1.2 Statistical Pattern Recognition

In the statistical methodology, data samples are represented in terms of d measurements and form a d-dimensional space called *feature space* [56]. The d-dimensional feature vectors are then used to build a classification system. The basic phases in the operation of a pattern recognition system are illustrated in Figure 1.1

The pattern recognition algorithm works in two main stages, named training and testing. In the training stage, some preprocessings are initially applied to the training samples. In the context of CAD, this preprocessing stage can include noise reduction, contrast enhancement, image registration and segmentation of the region of interest in the acquired image. The next stage is to extract/select some features from the images which properly highlight dissimilarities between the different classes (e.g. benign versus malignant). These features are then used in the next stage for constructing/training a classifier in order

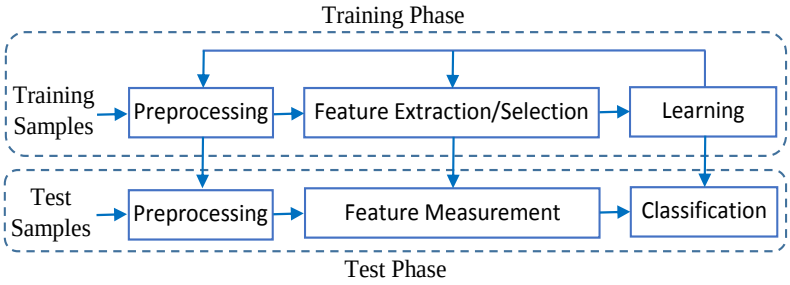


Figure 1.1: General framework for implementing a statistical pattern recognition system.

to assign each input pattern to one of the classes in the classification problem. The feedback path from the learning phase to the preprocessing and feature extraction modules enable a designer to modify the computed attributes.

The proposed CAD systems in this thesis concern *supervised learning* where a set $\{(x_1, \omega_1), \ldots, (x_D, \omega_D)\}$ is used for building a classification system. In this set, x_i is the feature vector of the ith pattern and ω_i is its class label taken from the group $\Omega = \{\omega_1, \ldots, \omega_C\}$ for a C-class problem.

In the testing stage, the same preprocessing and feature measurement operations are applied to a given unlabeled pattern and the classifier assigns it to one of the C classes.

The main focus of this thesis is on designing (new) feature extraction and classification algorithms for CAD purposes. In the following sections, the basic feature measurement and classification methods that are used throughout the thesis are reviewed.

1.3 Statistical Feature Extraction

Principal Component Analysis (PCA) [61] and Independent Component Analysis (ICA) [52] are two statistical methods that are used for

data modeling and feature extraction in this thesis. The reasons for adopting theses methods are bi-fold:

1. The effectiveness of PCA and ICA for feature extraction has been widely proven in different pattern recognition applications,
2. They belong to the family of the feature learning techniques and as data-driven methods can,
    - make learning algorithms less dependent on human prior knowledge and,
    - facilitate the procedure of extracting useful information when building classifiers [5].

1.3.1 Principal Component Analysis

PCA is a popular statistical approach for feature extraction, dimensionality reduction and data visualization [61]. Given a data set of random vectors $X = \{x_1, \ldots, x_D\}$ where $x_n \in \mathbb{R}^N$, the PCA algorithm gives a representation of the data in $\mathbb{R}^M (M < N)$ such that the new variables are less redundant compared to the original ones. PCA can be defined based on projection error minimization. In this formulation, each data vector can be exactly represented by a linear combination of a complete orthonormal set of *basis vectors* $\Phi = \{\phi_1, \ldots, \phi_N\}$ as,

$$x_n = \sum_{i=1}^{N} a_{ni}\phi_i \tag{1.1}$$

where the coefficient set $\{a_{n1}, \ldots, a_{nN}\}$ is unique for the data vector x_n. Each coefficient a_{ni} can be computed based on the orthonormality property and by projecting x_n onto the subspace spanned by the ith basis vector,

$$a_{ni} = x_n^T \phi_i. \tag{1.2}$$

The set of basis vectors, which are called *principal components* (PCs), can be obtained by computing the eigenvectors of the data

covariance matrix [61]. To represent a data vector x_n in a lower-dimensional subspace, one can retain the $M < N$ eigenvectors with the largest eigenvalues and approximate it by

$$x_n \approx \sum_{i=1}^{M} a_{ni}\phi_i. \tag{1.3}$$

The first step in building the PCA model is centering the data that can be done by first computing the mean vector of X and then subtracting every data vector from it. In the second step, the covariance matrix of the centralized data set is computed. The final step is to find the eigenvector matrix Φ and diagonal eigenvalue matrix Λ of the covariance matrix.

1.3.2 Independent Component Analysis

In the classic ICA model [52], it is assumed that an observed random vector $y \in \mathbb{R}^N$ is generated by linearly combining the elements of a constant matrix $B \in \mathbb{R}^{N\times N}$, named the *mixing matrix*, with a vector of random variables $s \in \mathbb{R}^N$:

$$y = Bs. \tag{1.4}$$

Since the rows of the mixing matrix B (also known as basis vectors) are fixed, elements of the vector s should change for any new data to generate it using (1.4). The aim of the ICA model is to estimate both B and s using a large enough set of observed random vectors. The linear model presented in (1.4) is based on three main assumptions [52]. The first one is that the components of s are statistically independent and they are called *independent components* (ICs). The second assumption is that the ICs have *nongaussian* distributions. Finally, it is assumed that the mixing matrix is invertible.

Data Preprocessing

Adopting a 2-step preprocessing strategy including centering and whitening allows an easy implementation of the ICA model. The first step is

done by removing the mean vector of y from it, which also implies that s is zero-mean. The aim of performing the second step is to linearly transform a zero-mean observed vector y into a vector x by a whitening matrix V such that the elements of x are uncorrelated and have unit variance. In this way, the effect of the second-order statistics would be eliminated. Then, the whitening transform is given by

$$x = Vy = VBs = As \tag{1.5}$$

where A is the new mixing matrix. An advantage of using a whitening algorithm is that it makes the mixing matrix orthogonal and less parameters for its estimation are required [52].

There are several whitening transformations in which PCA is one of the most popular choices. The whitening matrix is then given by

$$V = \Lambda^{-1/2}\Phi^T. \tag{1.6}$$

where Φ and Λ are eigenvector and eigenvalue matrices, respectively. By using PCA, one can also retain only the most important eigenvectors of data and by discarding the rest, considerably reduce the computational complexity of the ICA algorithm.

ICA by Maximization of Nongaussianity

The ICs can be obtained by finding the inverse of the mixing matrix which is called the *unmixing matrix* W. Estimating the jth IC is then equivalent to finding proper approximation of the jth row of W, denoted by w_j^T, and linearly combining the elements of x by it:

$$u_j = w_j^T x = \sum_i w_{ji} x_i. \tag{1.7}$$

An intuitive methodology for estimating the ICA model is to maximize the nongaussianity of the ICs. This approach is based on the relationship between nongaussianity and independence (see chapter 8 of [52]) and can be interpreted by a cornerstone theorem in probability theory known as the *Central Limit Theorem* (CLT). The CLT

says that, subject to certain conditions, the sum of infinite number of independent and identically distributed random variables, which is itself a random variable, tends to have a Gaussian distribution. As a consequence, the sum of multiple independent random variables is more likely to be Gaussian distributed than each of the variables alone. Thus, the CLT motivates to find a linear combination of ICs that are maximally nongaussian.

Negentropy as a Measure of Nongaussianity

Comon [22] proposed to use *negentropy* J for measuring the distance to gaussianity of a random vector u. The negentropy approach is based on the concept of the *differential entropy*. For a random vector $u \in \mathbb{R}^N$ with probability distribution function (PDF) $p(u)$, the differential entropy is given by

$$H(u) = -\int p(u) \log p(u) du. \tag{1.8}$$

A well-known proposition of information theory says that a Gaussian variable has the largest entropy between all random variables of equal variance. By taking advantage of this property, negentropy is defined as

$$J(u) = H(u_{gauss}) - H(u) \tag{1.9}$$

where u_{gauss} stands for a Gaussian random vector with the same covariance matrix as u.

Negentropy is always nonnegative and is equal to zero if and only if the PDF of u is Gaussian. Therefore, departure of a random variable from gaussianity can be determined by the quantity of its negentropy. For the random variable in (1.7), maximizing $J(u_j)$ leads to estimation of one of the ICs. This maximization could be done by utilizing an accurate approximation of the negentropy [51]:

$$J(u_j) \approx c\{\mathbb{E}[G(u_j)] - \mathbb{E}[G(v_j)]\}^2 \tag{1.10}$$

where c is a constant, G is a non-quadratic function and v_j is a Gaussian variable of zero mean and unit variance.

1.4 Feature Selection

Since the ultimate goal of constructing a pattern recognition algorithm is to accurately distinguish samples of different classes, a key issue in designing a classification system is that the employed features provide significant discriminatory information. The idea behind some of the feature extraction methods, however, is to model the whole set of data regardless of their class labels. Many of efficient feature extraction approaches (e.g. PCA and ICA) fall into this category and are called *unsupervised* methods because they do not consider patterns' labels in their algorithms. Moreover, some of the extracted features may be redundant measures of the patterns, which do not add new information to the classification problem, and their existence in the feature vector could degrade the performance of a classification system due to *overfitting*.

Therefore, incorporation of a feature selection approach that takes the discriminatory information provided by each feature into account could enhance the performance of a pattern classification system. In the pattern recognition literature, one can recognize two main groups of the feature selection methodologies namely *filter* and *wrapper* [46]. In the filter methods features are selected independent of the choice of the classifier and appropriateness of each feature for the classification task is evaluated by some simple statistics. Wrapper methods, on the other hand, make use of a classifier to determine the predictive powers of the features.

In this thesis the P-metric method [55], which is a filter algorithm, was used to select relevant features:

$$P(Feature_i) = \frac{|\mu_{i1} - \mu_{i2}|}{\sigma_{i1} + \sigma_{i2}} \tag{1.11}$$

where μ_{i1} and μ_{i2} are respectively the means of the samples belong to the first and second classes in the subspace of the ith computed feature and σ_{i1} and σ_{i2} are their corresponding standard deviations. A high P-metric value for a feature implies that it provides a good separation between the samples of the two classes.

1.5 Classification

The classification systems that are used in this thesis can be categorized into two broad groups, namely *single* and *ensemble* classifiers. While the major concern in designing a single classifier is to tune its parameters for having the best performance, an ensemble classification methodology employs a group of relatively sub-optimal classifiers and by fusing their decisions can yield optimal decision [64, 122].

1.5.1 Single Classification Systems

Classic K-Nearest Neighbor

In the classic *K-Nearest Neighbor* (KNN) method [24], a given test pattern with unknown label is assigned to the class with largest number of patterns in a neighborhood of K closest training patterns to this test sample (Figure 1.2). Several distance metrics can be used to measure similarity between a test and training patterns in which *Euclidean distance* is the most common one.

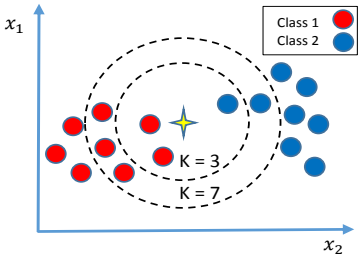


Figure 1.2: Example of classifying a test sample (yellow star) with unknown label in 2-D features space using KNN.

Distance-Weighted KNN

The *Distance-Weighted K-Nearest Neighbor* (DWKNN) method [34] is an extension to the classic KNN approach and is based on the rationale that a training pattern close to a test sample should have higher weight compared to another training observation which is at a greater distance from the test pattern (Figure 1.3). For a given test sample x, the set of Euclidean distances $(d_1, \ldots, d_K)$ that corresponds to the K-nearest training samples $(x_1, \ldots, x_K)$ is used to assign a weight w_i to the ith nearest neighbor x_i as follow:

$$w_i = \begin{cases} \frac{d_K - d_i}{d_K - d_1} & \text{if } d_K \neq d_1 \\ 1 & \text{if } d_K = d_1 \end{cases}. \tag{1.12}$$

In a C-class problem, the weight of class C_j is the sum of all weights assigned to its samples in the KNN set:

$$w'_{C_j} = \sum_{x_i \in C_j} w_i. \tag{1.13}$$

After normalizing the class weights as,

$$w_{C_j} = \frac{w'_{C_j}}{\sum_{i=1}^{C} w'_{C_i}} \tag{1.14}$$

the test sample x is assigned to the class with the greatest weight value. Note that, DWKNN is equivalent to KNN when $K = 1$.

Local Mean-Based

The Local Mean-Based (LMB) classification method [82] uses the KNN training samples of each class for decision making. Let $\{x_1^i, \ldots, x_D^i\}$ be the set of training samples belonging to class ω_i and let $\{x_1^i, \ldots, x_K^i\}$ be the subset of K-nearest samples from this class to a given test pattern

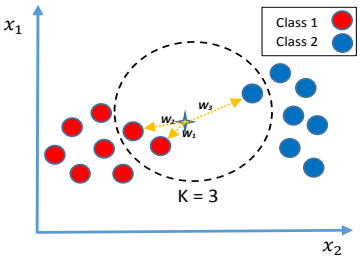


Figure 1.3: Example of classifying a test sample (yellow star) with unknown label in 2-D features space using DWKNN. A weight value is assigned to each of the train patterns in the KNN set which is proportional to the inverse of its distance to the test sample. In this example, $w_1 > w_2 > w_3$.

x. The local mean vector of class ω_i is computed by using the KNN set:

$$v_i = \frac{1}{K} \sum_{j=1}^{K} x_j^i. \tag{1.15}$$

In the LMB method, a test sample is assigned to the class with the closest mean vector (Figure 1.4). It has been shown that the LMB classification method is robust to outliers and has favorable performance in classification problems with high dimensionality and small training sample size [82].

Support Vector Machines

Support Vector Machines (SVM) works based on the idea of the maximum margin solution and finds a hyper-plane that has the greatest distance to the training samples at the boundaries of a binary classification problem [12, 117].

Let in a binary classification problem with D training patterns $\{x_1, \ldots, x_D\}$, labels being taken from the set $\{-1, 1\}$ where $\omega_i = 1$ and $\omega_i = -1$ show that x_i belongs to class 1 or class 2, respectively.

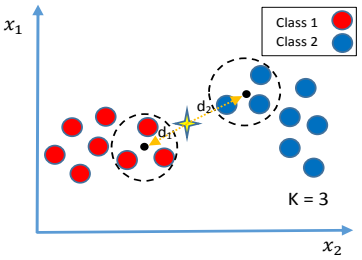


Figure 1.4: Example of classifying a test sample (yellow star) with unknown label in 2-D features space using LMB. Based on the distances between the test sample and local means of the classes with $K = 3$, the test sample is assigned to class 1 ($d_1 < d_2$).

Suppose there is a hyperplane that separates samples of the two classes (Figure 1.5). A set of patterns $\mathbf{x}$ that lie on the hyperplane satisfies:

$$\mathbf{w}.\mathbf{x} + b = 0 \tag{1.16}$$

where $\mathbf{w}$ is normal to the hyperplane, $\|\mathbf{w}\|$ is the Euclidean norm of $\mathbf{w}$ and $|b|/\|\mathbf{w}\|$ is the perpendicular distance from the hyperplane to the origin [12].

When the classes are linearly separable, the SVM algorithm looks for two hyperplanes with largest distance (margin) between them. These hyperplanes can be formulated as,

$$x_i.\mathbf{w} + b \geq +1 \;\; for \;\; \omega_i = +1 \tag{1.17}$$

$$x_i.\mathbf{w} + b \leq -1 \;\; for \;\; \omega_i = -1 \tag{1.18}$$

which can be combined and rewritten as,

$$\omega_i(x_i.\mathbf{w} + b) - 1 \geq 0 \;\; \forall i. \tag{1.19}$$

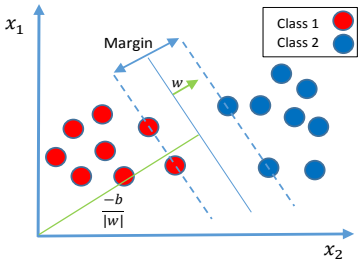


Figure 1.5: Separating hyperplanes in a linear classification problem computed by SVM. The support vectors are circled.

The SVM optimization problem can then be presented as,

$$\begin{cases} \mathbf{arg\ min}_{(\mathbf{w},b)} \|\mathbf{w}\| \\ subject\ to \\ \omega_i(x_i.\mathbf{w}+b)-1 \geq 0 \ \ \forall i \end{cases} \tag{1.20}$$

which gives the maximum perpendicular distance between the decision hyperplane and the closest point from the training set. The term $\|\mathbf{w}\|$ in (1.20), which involves a square root and is difficult to solve, can be substituted by $\frac{1}{2}\|\mathbf{w}\|^2$ to form a *quadratic programming optimization* problem. The above optimization problem can be solved by making use of *Lagrange multipliers* α_i, one multiplier for every constraint in (1.20), by the following Lagrangian function:

$$L = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{D} \alpha_i\{\omega_i(x_i.\mathbf{w}+b)-1\}. \tag{1.21}$$

Note that the SVM algorithm can also be used for solving non-linear classification problems by taking advantage of the *kernel trick* [117].

1.5.2 Ensemble Classification Systems

When a classification problem is too difficult for a single classification system to solve, which could be due to considerable overlap of the different classes' samples, complex decision boundaries or too little data to train the classifier, combining the outputs of some sub-optimal classifiers that complement their decision boundaries has been proven to be an efficient approach to get high classification result [64, 93, 122].

Diversity of the single classifiers that are used for building an ensemble system is the keystone to the success of the combined algorithm [65, 93]. It means that the single classifiers make different errors on different patterns and each of them is expert in classifying patterns belong to a sub-region of the feature space.

There are three main strategies for building an ensemble classification system with diverse single classifiers [93]. The first strategy uses different subsets of the training patterns for building the single classifiers. An example of such ensemble system is *bagging* [11] where the subsets of the training patterns are obtained in a 'random sampling with replacement' procedure. The second strategy to have diverse classifiers is to use different classification methodologies (e.g. KNN, SVM, neural network) or the same type of classifiers with different sets of training parameters. Finally, the third strategy to encourage diversity in a combined classification system is to use feature sets of different natures or different subsets of a given feature space for building the single classifiers.

The last mentioned strategy is the basis for constructing ensemble classification systems in this thesis. To shed light on the process of diversifying the single classifiers using this strategy, an example is shown in Figure 1.6. The figure can be interpreted in two ways according to the approaches used for building the ensemble system.

In the first scenario, Figures 1.6 (a), (b) and (c) can be seen as the results of three different feature extraction methods in a 4-class pattern recognition problem where each color represents the area occupied by one of the classes in the feature space. In each of Figures (a) and (b), a pair of classes are well-separated from each other which means that a

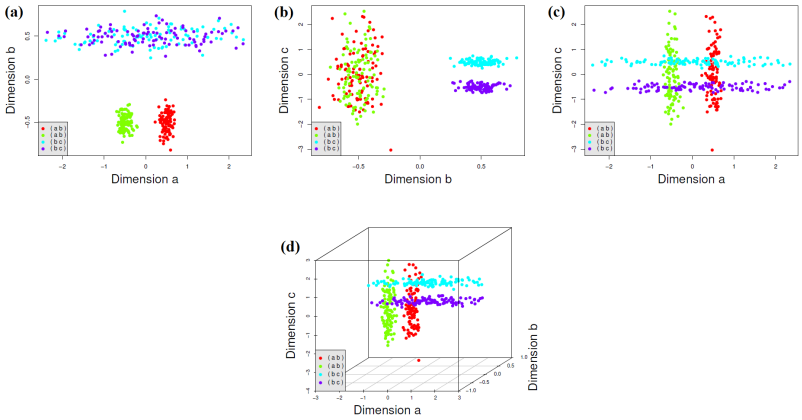


Figure 1.6: A 4-class pattern recognition problem where each color represents one of the classes [89]. Figures (a)-(c) can be seen as the results of different feature extraction approaches or sub-spaces taken from Figure (d).

classifier that is trained on this feature space can easily learn a decision boundary for categorizing samples belong to these two classes. However, the overlap between the samples of the other pair of the classes in each feature space is considerable and the classifier cannot yield high classification accuracy. Combining the outputs of these two classifiers may not improve the performance of each of them because when the first classifier makes a correct decision for the samples of one the class pairs, the second classifier can generate an uncertain or even wrong decision for the samples of the same pair of classes. In this situation, adding another classifier that is trained on a feature space like Figure 1.6 (c) to the ensemble system can complement the single classifiers' performances. Although the third classifier has a lower performance in categorizing the two pairs of classes compared to the first and second classifiers, it has a decent performance in categorizing samples of all classes. Therefore, one can expect that the outcome of combining these three classifiers would be satisfactory.

In the second scenario and from a different point of view, the feature spaces in Figures 1.6 (a), (b) and (c) can be considered as the

sub-spaces created from a higher-dimensional feature space shown in Figure 1.6 (d). In this case, the high overlap between the classes in the original feature space, which can be due to noisy data or some other measurements that can be irrelevant to the classification task, could be decreased by generating sub-spaces of lower dimension. In addition to improve class separability, this sub-space generation method has the important advantage of reducing the risk of overfitting which could happen when a feature space has a very high dimension and there is not a large set of data to properly train a classifier [9].

Diversity Measures

A group of measures have been defined to quantify diversity of a set of single classifiers in an ensemble system. These measures can be divided into two main groups namely *pairwise* and *non-pairwise* based on the number of classifiers that are employed in the measurement phase [65]. While a pairwise measure is designed for two classifiers, a non-pairwise approach takes the diversity of the whole ensemble system into account. In this thesis, a non-pairwise diversity measure called *interrater agreement* κ is used.

Let $P = \{P_1, ..., P_Q\}$ be a system of Q classifiers and let $\{x_1, \ldots, x_D\}$ be a set of data containing D labeled patterns. Let $O_i = [O_{i1}, ..., O_{iQ}]$ be the joint output of the ensemble system for the ith input sample where O_{ij} denotes the output of the jth single classifier and it is defined to be 1 if P_j categorizes x_i correctly and 0, otherwise. Let $q(x_i)$ denotes the number of classifiers from P that correctly categorize the input sample x_i expressed by

$$q(x_i) = \sum_{j=1}^{Q} O_{ij}. \tag{1.22}$$

Denote by

$$r_j = \frac{1}{D} \sum_{i=1}^{D} O_{ij} \tag{1.23}$$

the classification rate of the jth classifier, the average classification accuracy of the ensemble system is defined by

$$\overline{r} = \frac{1}{Q} \sum_{j=1}^{Q} r_j. \tag{1.24}$$

Using the notation presented above, the κ measure of diversity is given by

$$\kappa = 1 - \frac{\sum_{i=1}^{D} q(x_i)(Q - q(x_i))}{DQ(Q-1)\overline{r}(1-\overline{r})} \tag{1.25}$$

where $\kappa = 0$ indicates that the single classifiers are independent. Small values of κ leads to better diversity and a negative κ shows negative dependency (high diversity) among the classifiers [65].

Classifier Combination

After training a set of diverse classifiers, a strategy should be adopted to combine their outputs to have the final decision of the ensemble system. One possible categorization of the existing combining approaches is: class label combination vs. combining the continuous outputs [93]. In the first combination approach, a combination rule is applied to the class labels generated by the individual classifiers for a given test pattern. A widely used technique of this family of combination rules, which is employed in this thesis, is *majority voting*. In this technique, the number of votes (i.e. output labels) in favor of each class is counted and the test pattern is then simply assigned to the class that could achieve the highest vote.

In the second combination methodology, the output of an individual classifier for a given class is interpreted as the degree of support or an approximation of the posterior probability provided for that class. Algebraic rules are popular members of this family of combination techniques where a linear function $\xi(.)$ is used to merge the individual classifiers' outputs. Let for an ensemble system with Q classifiers,

$$\vartheta_j(x_i) = \xi(\{o_{1j}(x_i), \ldots, o_{Qj}(x_i)\}) \tag{1.26}$$

denote the result of applying the combination function to the jth output of the single classifiers corresponding to class ω_j for a given test pattern x_i. The common algebraic rules that work as the linear combination functions are *sum, product, minimum, maximum* and *median* rules [64, 93]. In this thesis the sum and product rules are used for classifier combination which can be presented respectively as,

$$\vartheta_j(x_i) = \frac{1}{Q}\sum_{i=1}^{Q} o_{ij}(x_i), \tag{1.27}$$

$$\vartheta_j(x_i) = \frac{1}{Q}\prod_{i=1}^{Q} o_{ij}(x_i). \tag{1.28}$$

In Chapter 3, another type of combination rule that is applied to continuous outputs is used. It is called *Dempster's rule* of combination [27] and is appropriate for dealing with problems which include uncertain information. Details of this combination rule along with the motivation for using it in a CAD problem is described in Chapter 3.

1.6 Performance Evaluation

It is common that the performance of a CAD system, which is employed to solve a binary classification problem, is evaluated through some statistical measures. These measures are *accuracy* (ACC), *sensitivity* (SE) and *specificity* (SP) and are defined as

$$\begin{cases} ACC = \frac{TP+TN}{TP+TN+FP+FN} \\ SE = \frac{TP}{TP+FN} \\ SP = \frac{TN}{TN+FP} \end{cases} \tag{1.29}$$

where

- True positive (TP): number of pathological patterns correctly classified as abnormal.
- True negative (TN): number of healthy patterns correctly classified as normal.
- False positive (FP): number of healthy patterns incorrectly classified as abnormal.
- False negative (FN): number of pathological patterns incorrectly classified as normal.

1.7 Ultrasound-Based CAD

The CAD algorithms presented in this thesis are based on *ultrasound imaging* [107]. This imaging modality has a broad range of usage in medical applications because it is a safe device which does not use harmful ionizing radiations, it provides clinicians with real-time images, it is portable and relatively cheap. However, the main disadvantage of the ultrasound images is their limited resolution and low signal to noise ratio (SNR) which can limit the usage of some image processing algorithms for quality enhancement. Another drawback of the ultrasound images is that their quality or relevance is dependent on the operator's skill. Table 1.1 summarizes the specifications of ultrasound as well three other widely-used image modalities.

1.8 Outline of the Thesis

The thesis is divided into two main parts. Part I is concerned with:

1. Designing a new CAD system for prostate cancer biopsy by focusing on handling uncertainties in labels of the ground truth data.
2. Using ICA for learning features from radiofrequency (RF) signals backscattered from prostate tissues.

Table 1.1: Comparison of the specifications of the different imaging modalities [107].

Modality	Ultrasound	X-ray	CT	MRI
What is imaged	Mechanical properties	Mean tissue absorption	Tissue absorption	Biochemistry (T_1 and T_2)
Access	Small windows adequate	2 sides needed	Circumferential around body	Circumferential around body
Spatial Resolution	Frequency and axially dependent	$0.3-3mm$	$\approx 1mm$	$\approx 1mm$
Penetration	Frequency dependent $3-25cm$	Excellent	Excellent	Excellent
Safety	Very Good	Ionizing radiation	Ionizing radiation	Very Good
Speed	100 frames/sec	Minutes	0.5 minute to minutes	10 frames/sec
Cost	Low	Low	High	Very High
Portability	Excellent	Good	Poor	Poor

3. Proposing a new ensemble scheme for learning ICA dictionaries from RF signals backscattered from a tissue mimicking phantom.

In Chapter 2, prostate anatomy is briefly described and the procedures of prostate cancer diagnosis and computer-aided prostate biopsy based on ultrasound are reviewed.

The proposed method for dealing with uncertainties pertaining labels of the data in computer-aided biopsy of the prostate is presented in Chapter 3. The label uncertainties are tackled in the framework of the theory of evidence and the size of the training data is increased by assigning soft labels to the group of biopsy samples with uncertain labels and including them in the training phase of the classifier.

The idea of using ICA for learning features from medical RF signals is presented in Chapter 4. It is demonstrated that ICA features can provide better classification results on the prostate biopsy samples than the traditional methods devised for extracting features from the RF signals. Furthermore, an ensemble scheme for learning ICA dictionaries from the RF signals is also proposed in this chapter.

In Part II of this thesis, the issue of detecting myocardial infarction (MI) is addressed as outlined below:

1. Statistical modeling of the temporal deformation patterns of the left ventricle (LV) to detect abnormalities in its regional function.

2. Statistical modeling of the spatio-temporal patterns of LV function for detecting MI.

3. Statistical modeling of the LV shapes for detecting MI.

In Chapter 5, heart anatomy and the cardiac cycle are briefly described; principles of myocardial deformation estimation are concisely reviewed and the main myocardial deformation imaging techniques are introduced.

Chapter 6 introduces an automatic method for detecting abnormalities in regional myocardial deformation. The CAD system uses PCA to model the whole temporal profile of the strain (rate) curves. The obtained classification results show that the PCA-derived features can provide better description of the deformation traces compared to the common traditional techno-markers.

In Chapter 7, the method presented in Chapter 6 is moved one step further and PCA is used to provide a spatio-temporal representation of the LV deformation. It is demonstrated that the proposed representation is more efficient than the widely-used traditional features and it is able to provide better results than an expert cardiologist.

A PCA-based statistical shape model for the LV is presented in Chapter 8. The LV shapes are divided into non-overlapping segments and independent PCA models are built on them. From each model, a small subset of PCs that provides relevant discriminatory information is selected in a feature selection phase. The selected features are then connected together to form a global representation of the LV shape. Experimental results show that the proposed local-to-global shape vector can better differentiate normal LV shapes from the infarcted ones as compared to a shape vector derived by modeling the global LV shapes with PCA. It is also demonstrated that there are direct links between some of the first selected modes of variation of the global PCA model and patho-physiology of MI.

Finally, Chapter 9 summarizes the thesis, draws a general conclusion and outlines some perspectives for future research.

Part I

Computer-Aided Prostate Cancer Biopsy

Chapter 2

Ultrasound-Guided Prostate Biopsy

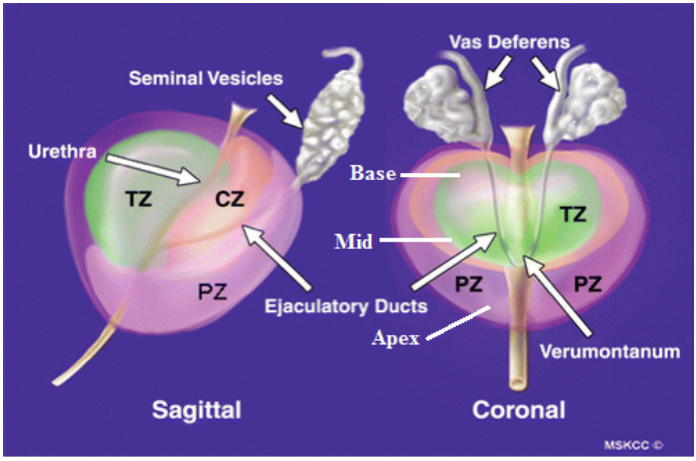


Figure 2.1: Prostate anatomy. [26].

2.1 Prostate Anatomy

The prostate is a chestnut shaped gland located between the pelvic bones, below the bladder and in front of the rectum. This relatively small organ, which has the average weight of 11 grams for healthy adults, is part of the male reproductive system.

The prostate gland is divided into base, mid an apex where apex refers to the upper part of the prostate close to the bladder. Based on the work of McNeal [80], it can also be partitioned into three zones namely the *transition zone* (TZ), the *central zone* (CZ) and the *peripheral zone* (PZ) (Figure 2.1). Two ejaculatory ducts and the urethra, which respectively come from the seminal vesicles and the bladder, pass through the prostate gland.

The TZ zone encompasses the proximal urethra, contains around 5% of the prostate tissue and approximately 10-20% of the cancers originate in this zone [81]. The CZ surrounds the ejaculatory ducts at the base of the prostate, comprises about 25% of the prostate tissue but only 2.5% of the prostatic cancers originate in this portion of the

gland [81]. Finally, the PZ surrounds up to 70 % of the prostate gland, comprises the posterior and lateral sides of the prostate and accounts for roughly 70-80% of prostate cancers [81].

2.2 Prostate Cancer Diagnostic Procedure

Prostate Cancer (PCa) is the most common cancer type among men and remains the second leading cause of death in men due to cancer. PCa can be diagnosed in its early stages with the use of a screening process. Figure 2.2 represents the current diagnostic procedure for PCa.

A blood test called *prostate specific antigen* (PSA) is usually the first step in the PCa screening procedure. PSA is a glycoprotein enzyme which is secreted by the epithelial cells of the prostate gland. The PSA test, measures the concentration of this enzyme in the blood. The PSA level in men with healthy prostates is low but is often increased due to PCa or other prostate abnormalities [15]. When the PSA level is increased to 3-4 nanograms per milliliter (ng/mL), patients are referred for further examination.

Another common diagnostic procedure to detect PCa is digital rectal examination (DRE). To do DRE, the physician inserts a lubricated, gloved finger into the rectum to inspect the prostate surface. Since a healthy prostate consists of soft and smooth tissue, a firm or stony feeling of the tissue is a sign of malignant prostate.

If the results of the PSA and/or DRE examinations provide some evidence of malignancy, usually a *trans-rectal ultrasound* (TRUS) guided biopsy is performed where ultrasound imaging is used to guide and monitor the biopsy needle in the different zones of the prostate gland. TRUS biopsy protocols are normally based on systematic sampling of those prostate zones that are more suspected of being cancerous.

As shown in Figure 2.2, the last stage of the diagnostic procedure is *histopathologic analysis* of the prostate tissue samples. In this stage, a pathologist evaluates the specimens and in the case of observing cancerous cells based on their microscopic appearance, a Gleason score

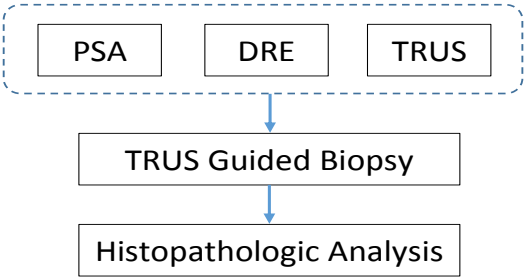


Figure 2.2: Prostate cancer diagnosis procedure. TRUS biopsy is normally performed if the results of the PSA and/or DRE examinations provide some evidence of malignancy.

[45] ranging from one to five is assigned to a sample where higher scores are associated to more aggressive cancers.

2.3 TRUS-Based Computer-Aided Biopsy

Performing a TRUS guided biopsy is traditionally based on a sampling protocol. The sextant protocol [50] is the first proposed sampling technique which involves six biopsies taken from the base, mid and apex of the prostate. A drawback of this technique, however, is that it under-samples the PZ and can miss some tumors there. Therefore, some modified sextant protocols were proposed to enhance diagnostic accuracy [94].

Prostate biopsy, however, is a an invasive process which can result in side-effects like bleeding, discomfort, urinary retention and infection [21] and in case of repeated or systematic sampling protocols involving many cores, complications could be more common.

As an alternative to the traditional biopsy protocols, a *computer-aided biopsy* (CAB) technique can be used to enhance the accuracy of the tissue sampling procedure. The idea is to provide physicians with a risk map superimposed to the B-mode image (Figure 2.3) in

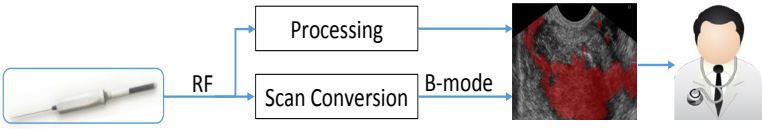


Figure 2.3: Implementation of a CAB of prostate using TRUS technique. RF or B-mode images are processed to detect risky regions in the prostate gland. The suspicious regions are then highlighted in the B-mode image to guide physician towards insertion of biopsy needle.

order to guide the biopsy needle into the suspicious regions. In this way, a lesion directed biopsy can be performed and the number of unnecessary biopsy cores can be decreased. To create the risk map, a machine learning algorithm could be trained with a set of labeled RF or B-mode images. Labels of the images are assigned to them by an expert physician as part of the histopathologic analysis of the prostate tissue samples.

In order to improve the accuracy of the biopsy procedure, ultrasound images can also be merged with the MR images to create the risk map [76]. In this way, the biopsy protocol takes advantage of both the detailed MR images and real-time ultrasound guidance.

Chapter 3

Soft-label Reinforced real-time CAB for Guided Prostate Tissue Sampling

This chapter is based on a paper published in *IEEE Int. Ultrason. Symp. 2013*, M. Tabassian, F. Galluzzo, L. De Marchi, N. Speciale, G. Masetti, and N. Testoni, "Soft-label reinforced rtCAB for guided prostate tissue sampling", pp. 880–883.

3.1 Abstract

In this study a real-time computer-aided biopsy (rtCAB) system is presented to support prostate cancer diagnosis. Different types of features are extracted from trans-rectal ultrasound data and an ensemble learning algorithm is used in the classification phase. A new label assignment method is also employed to provide soft or crisp class labels for unlabeled data. The proposed model can be implemented in a parallel fashion in order to provide real-time support to physicians during biopsy. Experiments on ground truth images from biopsy findings demonstrate that the proposed approach can properly deal with unlabeled data and is able to provide better results than some examined supervised and semi-supervised classifiers.

3.2 Introduction

PCa is the second most frequent neoplasy in men, with almost one million new cases/year worldwide. Since PCa is not clearly detectable with standard, non-invasive diagnostic methods, TRUS guided prostate biopsy is mandatory. The procedure consists in a systematic tissue sampling by means of a surgical needle [94]. As patient discomfort and adverse event probability grow with each extracted sample, it is desirable to employ a CAB system to reduce the number of biopsy samples without negative impinging on diagnostic accuracy.

Satisfactory performance of such a system is highly dependent on proper features extracted from the data. Like other medical applications, the success of a CAB system depends on correct labels assigned to the data by domain experts. These systems may suffer from the uncertainty in the labels of the ground truth data due to the unknown tumor distribution in each bioptic sample. It is also important to select a suitable algorithm for learning characteristics of the data.

The mentioned above issues are considered in the current work to enhance the performance of a real-time computer-aided biopsy (rtCAB) system proposed in [112]. This system has been designed to

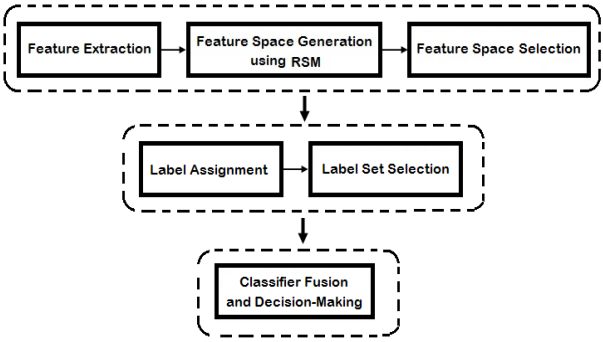


Figure 3.1: Architecture of the proposed ensemble classification system.

guide tissue sampling to areas with the highest probability of PCa and to provide real-time assistance to the physician during biopsy by implementation with the CUDA parallel processing platform. Since the effectiveness of using different types of features extracted from the TRUS data has been shown in several studies [74, 98], in this research features of different natures are employed in the structure of an ensemble classification model and partial labels assigned to some of the data are reconsidered by a label assignment phase. The label assignment method provides soft or crisp class labels for patterns with uncertain labels and in this way the size of the training set can be enlarged.

3.3 The Proposed Method

In this study, an ensemble algorithm is proposed to classify the TRUS data. For a large percentage of the data, the knowledge about the tumor volume of each tissue sample (core) provided by the physician is not enough to assign with certainty a given region of interest (ROI) of the core to a healthy or an unhealthy class. Therefore, a label assignment phase is employed to provide crisp or soft labels for the ROIs belonging to these cores. Steps of implementing the proposed method are shown in Figure 3.1 and each phase is described below.

3.3.1 Generating a Feature Space Pool

Inspired by the *Random Subspace Method* (RSM) [63], a pool of feature spaces is built to construct an ensemble network. Given a data set of random vectors $X = \{x_1, \ldots, x_D\}$ where $x_n \in \mathbb{R}^N$, different subsets of features of size $M < N$ are randomly sampled from the original feature space with replacement and multiple classifiers are constructed in the low-dimensional feature spaces. Note that this strategy falls into the third category mentioned in section 1.5.2 for creating an ensemble system.

3.3.2 Classifier/Feature Space Selection

By training a classifier on each of the feature spaces generated by random sampling, a pool of classifiers will be obtained. One of the necessary conditions for the success of the proposed method, as well as other ensemble networks, is the appropriate choice of the base classifiers. In order to choose an optimal subset of the base classifiers from the classifier pool, the *forward search* [46] algorithm, known as the most intuitive greedy approach, is used to explore the classifier pool and the *majority voting error* [96] is incorporated to assess the quality of the selected classifiers. The forward search algorithm starts with the most accurate classifier. In a sequential manner and by examining the performance of the ensemble networks made by the first classifier and any of the remaining classifiers, a classifier that leads to the best value based on the majority voting error criterion is selected as the second classifier. This process is repeated and finally a subset of the best classifiers is chosen.

3.3.3 Label Assignment

In order to assign labels to the uncertain data in each of the selected feature spaces the approach proposed in [109] is employed. Let $\Omega = \{\omega_1, \ldots, \omega_C\}$ be a set of C classes and x_i be an unlabeled pattern described by N features. Let P be a $C \times N$ matrix containing prototype vectors of the main classes and let $DT_i = \{d_{i1}, \ldots, d_{iC}\}$ be the

set of distances between an unlabeled sample x_i and the C prototypes according to some distance measure. Label assignment for x_i is performed in a 3-step procedure:

Step 1: The minimum distance between x_i and the class prototypes is taken from the vector DT_i,

$$d_{min} = min(d_{il}) \;\; l = 1, \ldots, C. \tag{3.1}$$

Step 2: A value $0 < \mu < 1$ is calculated for each of the C classes using the following function:

$$\mu_l(x_i) = \frac{d_{min} + \beta}{d_{il} + \beta}, \;\; l = 1, \ldots, C \tag{3.2}$$

where $0 < \beta < 1$ is a small constant value that ensures the utilized function allocates a value greater than zero to each of the C classes even if $d_{min} = 0$.

Step 3: A value $0 < \tau < 1$ is defined and based on the level of ambiguity regarding the class membership of the unlabeled sample x_i, this sample could be assigned to a set of classes (*soft labeling*) or just one main class (*crisp labeling*).

In this method, close distances between a given pattern and the class prototypes is considered as an indication of ambiguity in the patterns label and in such cases a soft label is assigned to it. This way of assigning labels to the uncertain data has the following advantages over the classic semi-supervised methods:

- the procedure of label assignment is very fast and is performed in one step,
- when there is a huge amount of uncertainty concerning the class membership of an unlabeled sample, instead of removing it from the set of added samples or accepting it with a high level of label uncertainty, this sample is included in the set of training data with a soft label.

3.3.4 Choice of the Classifier

It is highly probable that after adding the unlabeled data to the training set, the classes do not have equal number of training patterns. The LMB method [82], is used as base classifier in the structure of the proposed ensemble scheme to face this problem. In this method, for a given test sample x_i the mean vector of class ω_j is computed using the K nearest training samples of this class to x_i. The test sample is then assigned to the class with the closest mean vector. By using the LMB method an equal number of samples from the classes are employed for computing the mean vectors and in this way the classification algorithm can deal with the unbalanced data more appropriately.

Because in the proposed ensemble system continuous (probabilistic) outputs of the classifiers are required, distances between a given sample x_i and the mean vectors of the classes are transferred to probability values as follows:

$$Pr_j = \frac{1/d_{ij}}{\sum_{j=1}^{C} 1/d_{ij}} \tag{3.3}$$

where Pr_j is proportional to the inverse distance between the pattern and ω_j and the term in the denominator is used to normalize the probability values.

The LMB approach is also used in the label assignment phase of the proposed method for computing the prototype vectors of the classes. An example of label assignment using the LMB method is illustrated in Figure 3.2.

3.3.5 Label Set Selection

In the label assignment phase, $K \times \tau$ different sets of labels are assigned to the unlabeled data in each of the selected feature spaces where K is the number of nearest samples taken from each of the main classes for computing its prototype. In a 2-step filtering phase, a subset of these label sets are selected.

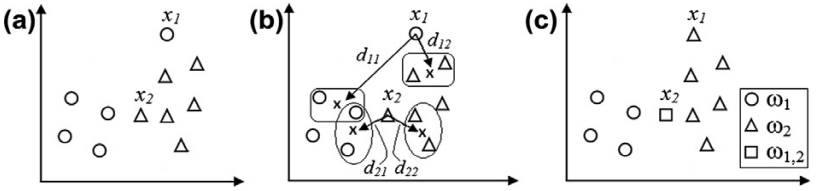


Figure 3.2: An example of the label assignment procedure using the LMB method with $K = 2$ [109]. (a) Training samples before label assignment, (b) computed class prototypes for samples x_1 and x_2, (c) training samples after label assignment. Since x_1 is closer to samples of ω_2 than those of ω_1 it is assigned to the second class. On the other hand, x_2 is close the borders of the classes and as a consequence, a soft label comprising both classes is assigned to this pattern.

In the first step, a threshold value TH is used to select only those label sets with the percentage of soft labels less than or equal to TH. Using this technique, the label sets that have a large number of soft labels and could not provide enough discriminatory information are discarded. In the second step, a label set selection algorithm is adopted to choose only one set of the labels for each selected feature space. Here, the forward search is used to explore among the classifiers. The performances of the classifiers are then considered as the label set selection criterion.

3.3.6 Classifier Fusion

Since soft (i.e. uncertain) class labels are assigned to some of the unlabeled data, the *Dempster-Shafer (D-S) theory of evidence* [27], which is a well-suited framework for reasoning with partial information, is used for classifier fusion and decision-making. The basic concepts of the D-S framework are reviewed below.

Let $\Omega = \{\omega_1, \ldots, \omega_C\}$ be a finite set of mutually exclusive and exhaustive hypotheses called the frame of discernment. A *basic belief assignment* (BBA) or *mass function* is a function $m : (2^{\Omega}) \rightarrow [0, 1]$

which satisfies the following conditions:

$$m(\emptyset) = 0 \tag{3.4}$$

$$\sum_{A \subseteq \Omega} m(A) = 1 \tag{3.5}$$

where $\emptyset$ is the empty set and a BBA that satisfies the condition (3.4) is called normal. The subsets A of Ω with non-zero masses are called the focal elements of m and $m(A)$ indicates the degree of belief that is assigned to the exact set of A and not to any of its subsets.

There are also two other definitions in the theory of evidence. They are *belief* and *plausibility* functions associated with a BBA and are defined respectively, as follow:

$$Bel(A) = \sum_{B \subseteq A} m(B) \tag{3.6}$$

$$Pl(A) = \sum_{A \cap B \neq \emptyset} m(B). \tag{3.7}$$

$Bel(A)$ represents the total amount of probability that is allocated to A, while $Pl(A)$ can be interpreted as the maximum amount of support that could be given to A.

Let m_1 and m_2 be two BBAs induced by two independent items of evidence. These pieces of evidence can be combined using Dempster's rule of combination which is defined as:

$$m(H) = \frac{\sum\limits_{A \cap B = H} m_1(A) \times m_2(B)}{1 - \sum\limits_{A \cap B = \emptyset} m_1(A) \times m_2(B)}. \tag{3.8}$$

After combining all pieces of evidence, a decision has to be made using the final belief structure. In this study, the *pignistic transformation*

[103] is used to drive probability functions from the belief functions. By uniformly distributing the mass of belief $m(A)$ among its elements for all $A \subseteq \Omega$, a pignistic probability distribution is defined as:

$$BetP(\omega) = \sum_{\{A \subseteq \Omega, \omega \in A\}} \frac{1}{|A|} \cdot \frac{m(A)}{1 - m(\emptyset)}, \ \forall \omega \in \Omega \tag{3.9}$$

where $|A|$ is the cardinality of subset A.

3.4 Experiments and Results

3.4.1 Ground Truth Data

We used a data set of TRUS video-sequences [112] collected in collaboration with the Department of Urology of S. Orsola Hospital in Bologna, Italy. The data set consisted of 2042 video sequences of 124 healthy and 84 unhealthy patients. RF echo signals were recorded using a TECHNOS (Esaote s.p.a.) ultrasonographer equipped with a trans-rectal probe EC123 employed with the maximum frequency of 7.5MHz and combined with a hardware and software platform for ultrasound signal acquisition [97]. Interested readers are referred to [40] for more details about signal acquisition and data preparation.

Tumor percentages of the cores, provided by an expert physician as part of the histopathological analysis, were used for categorizing them into labeled and unlabeled. We considered cores with tumor percentages greater than 90% as unhealthy class, benign cores as healthy class and ROIs belonging to cores with tumor percentages less than 90% as unlabeled data. The number of ROIs for the healthy class was 420 and an equal number of ROIs was selected for the unhealthy class. We used 2000 ROIs as unlabeled data in our experiments. The testing set was composed of 40% of the labeled data and the experiments were carried out using a 10-fold cross validation technique.

3.4.2 Experimental Settings

The original feature space was composed of 12 features including nine Unser features [116], one feature from the Nakagami distribution [102] and two features derived from the Von Mises distribution [75]. Subspaces of size five to nine were generated from the original feature space and based on the performances of the selected classifiers on the validation data, 7-dimensional subspaces were used in the proposed method. Size of the feature space pool was fixed at 50 and 11 feature spaces were selected to be used in the proposed ensemble model. In the label assignment phase, 10 values of K and 14 different τ were respectively examined from the sets $(1, 2, \ldots, 10)$ and $(0.3, 0.35, \ldots, 0.95)$. The value of TH in the label set selection phase was fixed to 55%.

3.4.3 Performance Comparison

Performance of the proposed method was compared to four single classifiers trained on the original 12-D feature space:

- A classifier based on the generalized discriminant analysis (GDA) [4] and Fisher linear discriminant [9] (FLD) which was also used in [112],
- SVM with RBF kernel,
- Multilayer perceptron (MLP) neural network and,
- The LMB method.

The MLP neural network was trained 10 times with random initializations by the Levenberg-Marquardt algorithm. It used default parameters, 15 epochs of training and had one hidden layer.

We also compared our method with two ensemble models. The first one was a supervised scheme that employed the same classifiers as those used in our system. It only trained on the labeled data and used the majority voting method for combining the outputs of its classifiers. The second one was a semi-supervised ensemble method named RASCO

Table 3.1: Average Classification Results (%) of the employed single and ensemble classification systems for different percentages of the training data.

	Percentage of the Training Data				
	10	20	30	40	50
Single Classifiers					
GDA+FLD	56.93	59.79	59.05	59.79	59.46
SVM	56.43	57.53	58.63	59.52	59.82
MLP	54.81	55.78	56.35	57	58.11
LMB	56.13	56.34	57.26	58	58.06
Ensemble Classifiers					
RSM with the Selected Classifiers	57.44	58.15	59	59.76	60.21
RASCO (11 Classifiers)	58.15	58.21	58.18	58.30	58.33
RASCO (12 Classifiers)	57.53	58.36	58.27	58.27	58.27
The Proposed Method	58.27	58.45	60.18	60.03	60.89

[60] which used also the unlabeled data in its training phase. This method is based on the RSM and employs the well-known Co-training algorithm [10].

3.4.4 Results and Discussion

Average classification results of the employed classification methods for different sizes of the training data are listed in Table 3.1. The range of examined KNNs for the classifiers based on LMB was $\{1, 2, \ldots, 20\}$ and the best test results of the classifiers are shown.

Clearly the overall performance of the proposed method is better than other listed classifiers and it is able to provide higher classification results than GDA+FLD which was used in the initial structure of the rtCAB system [112].

It can be seen that the first ensemble system in Table 3.1 (RSM with the selected classifiers) could enhance the performances of the single classifiers in most cases. It means that the employed subspaces and their corresponding classifiers complement each other. Incorporation

of these subspaces and also taking the uncertainty of the patterns' labels into account, enabled the proposed method to deal with the classification problem efficiently.

By comparing the performance of the proposed method with that of RASCO as a semi-supervised approach, it can be concluded that utilizing the proposed label assignment method and decision-making in the D-S framework is a proper way for handling uncertainties of the data labels.

3.5 Conclusion

In this study, a method for enhancing the performance of the rtCAB system presented in [112] is proposed by considering uncertainties in the labels of the ground truth data. Labels of the data were assigned by an expert physician, providing tissue description and illness percentage for each core but these partial labels could not be employed for learning a supervised classifier system. By making use of a label assignment phase, soft or crisp class labels were assigned to the unlabeled patterns. In this way, size of the training data was increased and collected data wasting avoided. Features of different nature were extracted from the TRUS data and an ensemble classification system was constructed by exploiting diverse subsets of them. The D-S theory of evidence was used to manipulate the accepted uncertainties in the data and to aggregate decisions of the base classifiers. The proposed method could be implemented using the CUDA parallel processing platform to provide an on-line support for physician during biopsy.

Chapter 4

Learning Features from Medical Radiofrequency Ultrasonic Signals

This chapter is based on: 1) a paper published in *IEEE 27th Int. Symp. Comput. Med. Syst. 2014*, M. Tabassian, N. Testoni, L. De Marchi, F. Galluzzo, N. Speciale and G. Masetti, "Learning Features from Medical Radiofrequency Ultrasonic Signals by Independent Component Analysis", pp. 62–65 and, 2) a paper which is under preparation for submission to a journal.

4.1 Abstract

This study proposes to learn features from medical radiofrequency (RF) ultrasonic signals by ICA. The reason for adopting ICA as a dictionary learning method is that such a data-driven technique is able to bring the important advantage of high generalization power over the conventional methods tailored for extracting features from the medical RF signals. While the conventional feature extractors suffer from their limiting assumptions about the complex interaction between ultrasound wave and tissue, ICA can learn higher-order statistical structures of a set of RF signals from the data set itself.

A new ensemble scheme based on the ICA dictionaries is also proposed for the classification of the RF signals. The main idea is to learn several dictionaries of features from the RF signals by ICA and employ a subset of them, that provides complementary descriptions of the data, in the structure of an ensemble classification system.

Experiments on clinical and a tissue mimicking phantom data demonstrate the superiority of ICA as compared to the conventional techniques designed for extracting features from RF echo signals. It is also demonstrated that the proposed ensemble ICA scheme is able to enhance the performance of the classic ICA model and a recently proposed ensemble ICA model.

4.2 Introduction

Medical ultrasound has a broad utility in examining tissue structures. This is because of the advantages of ultrasonic over other imaging modalities including patient safety, real-time performance, portability and economic cost of equipment. The success of the ultrasound-based clinical tools in characterizing tissue properties is heavily dependent on suitable features extracted from echo signals. The conventional methods tailored for extracting features from the backscattered echos are based on some presumptions about the interaction between ultrasonic waves and tissue. However, this interaction is very complex and there

are also other sources of complication introduced by instrument settings and signal acquisition conditions. Therefore, the conventional methods may fail to capture essential parameters of the backscattered signals.

An alternative approach for computing features from the ultrasound echos is to use feature learning strategies [1, 49, 52] which have attracted the interests of many researchers recently. The idea behind these methods is that for extracting appropriate features from a set of data, the data set itself can be incorporated into a learning scheme. This new family of feature extractors includes techniques that make few general assumptions about the data and exploit their higher-order statistical structures with unsupervised learning strategies. The goal of the learning phase is to find an efficient representation of the data by making use of a dictionary of basis functions and a weight vector.

In the literature of medical ultrasound, there are a few works that employed feature learning algorithms for tissue characterization. In [32], Independent Component Analysis (ICA) was used for extracting features from B-mode images and it was shown that this learning approach can outperform traditional feature extraction methods in characterizing numerical simulated tissues. The application of the texton-based feature learning technique [118] for the evaluation of cancer treatment effects was presented in [41]. A quantitative ultrasound spectral parametric map was built using the signals acquired from animal tumors and it was shown that the learned features from this map enable the classifier to provide very good discrimination between pre- and post-treatment states.

In this study, the ICA-derived features are used for the classification of medical radiofrequency (RF) ultrasonic signals. A new ensemble scheme based on ICA dictionaries is also proposed to improve the performance of the classic ICA method. Although advantages of the features learned by ICA have been proven in different research areas including medical and brain imaging [13, 32], computer vision [3] and speech recognition [66], finding a set of efficient ICA parameters is known to be a challenging task. The proposed ensemble system seeks to collect such an efficient set of parameters that can properly model

the data. It is also designed to provide favorable classification results using the learned ICA parameters where it has been demonstrated that by using ensemble classification frameworks, considerable performance improvements on the single classifier models can be achieved [64, 67]. One of the favorable properties of an ensemble scheme is that it allows to use multiple feature spaces in its structure. This advantage motivated us to incorporate a set of ICA dictionaries that provide complementary descriptions of the RF signals into an ensemble classification framework. Experimental results verified that the proposed ensemble model is able to outperform a single classifier that uses only one ICA dictionary to categorize the RF signals. Note that, the reason for learning features from the RF echos rather than from B-mode images is the richer information content of these signals.

4.3 Conventional Feature Extractors

The conventional methods proposed for extracting features from medical RF echos can be divided into four main categories:

4.3.1 Spectral

These methods are based on the idea that frequency-dependent properties of tissue-scatterers can be estimated by studying the spectrum of the backscattered RF echos. The extracted parameters from the spectrum convey valuable information regarding changes in tissue microstructure and can be used by a classifier to categorize normal and diseased tissues.

Lizzi et al. [70–72] proposed a well-established method for analyzing the relation between spectral parameters and attributes of tissue microstructure. First, an average power spectrum of the RF backscattered signals inside a ROI is computed. This spectrum is then calibrated to eliminate undesirable system effects. The authors realized that the obtained spectra usually exhibit quasi-linear shapes and suggested to fit a regression line to each of them. Three spectral parameters that have relation to physical scatterers properties of tissue are

then computed using the regression line. They are spectral intercept, spectral slope and midband fit.

Several studies have also demonstrated the usefulness of the frequency-domain parameters. It has been shown in [119] that variations of a spectral parameter, called integrated backscatter (IB), can be employed to differentiate between normal and myopathic myocardium. In an in vitro study of kidney [54], normalized power spectral density of echos has been used to estimate backscatter coefficients, scatterers' sizes and scattering strength. In [68], the spectral features proposed in [70] have been employed in a tissue characterization technique to discriminate between coronary plaques from asymptomatic and symptomatic patients. An in vivo plaque characterization method based on spectral analysis of backscattered intravascular ultrasound (IVUS) data has been presented in [88] to classify four types of coronary plaques.

4.3.2 Statistical

The rationale behind this family of feature extractors is that the statistical features of the backscattered echos provide useful tissue discrimination information. Since backscattered signals from tissue have a stochastic nature, one can select a PDF to capture statistical properties of the echos.

Early studies suggested to model the statistics of the envelope of backscattered echos from tissue by the Rayleigh distribution [115, 120]. It has been shown that the Rayleigh distribution can properly describe the statistics of the so-called fully-developed scattering process where the number of scatterers in the resolution cell is very large. However, this distribution fails to model more complicated structures of tissues where the scatterers are not necessarily fully-developed or contain structural components. This was the motivation for utilizing the K distribution, as a more general statistical technique than the Rayleigh model, for characterizing the statistics of the backscattered envelope [35, 86, 101]. In [6] the authors have shown that the K distribution is also a suitable method for modeling the raw RF signals and employed

this distribution to describe the statistics of the RF signals backscattered from the myocardium.

Although the K distribution is a well-suited framework for modeling different types of tissues, its application is limited because of the intensive computations required for approximating its parameters. Two main properties of the Nakagami distribution, namely generality and light computational load, have captured the interests of the researchers to employ it in the medical ultrasonic field with the aim of modeling statistics of the backscattered envelope [85, 102]. Another inclusive, yet computationally feasible technique for modeling statistics of the RF echo signals is the Generalized Gaussian Distribution (GGD) that has been proposed by Bernard et al. [7]. The authors showed that this distribution can describe statistics of a wide range of scattering conditions in medical ultrasound thanks to its ability to represent different distributions. It has been demonstrated that the GGD can model fully-developed speckle of the blood regions and partially-developed speckle of the myocardial regions with the Gaussian and heavy tail Laplacian-like distributions, respectively.

4.3.3 Time-Frequency

This class of feature extraction methods considers the interaction of the ultrasound wave with the scatterers in time and the effects of this interaction on the spectrum of the backscattered echo.

Flexibility and suitable localization in both temporal and spectral domains, make the wavelet transform as the most appealing method for performing time-frequency characterization of medical ultrasonic signals [18, 43, 44, 77, 78]. In [43], the continuous wavelet transform has been used to decompose an RF echo into its coherent and diffuse components which are associated to the large resolvable and small unresolvable tissue-scatterers, respectively. The proposed decomposition technique is based on the idea that the wavelet power has a large amplitude in the time location of a coherent component and by adopting a proper threshold value, the coherent components can be distinguished from the diffuse ones. It has been shown in [44] that the features ex-

tracted by this wavelet-based decomposition approach are useful for breast tissue characterization. A wavelet packet filtering method was proposed in [18] to decompose backscattered RF signals. A RGB false-colored image was generated from the set of sub-band images and it was discussed that this colored image is diagnostically helpful. Masotti et al. [77, 78] introduced a method based on the Discrete Wavelet Packet Transform (DWPT) for differentiating pathological regions. They performed a 3-level decomposition of the echos and analysed the results by a polynomial of degree four. The polynomial coefficients were then used to create a local histogram and the lower-order statistical properties of this histogram were considered as features and used for breast and prostate tissue characterization.

Moradi et al. [87] have shown that by continuously transmitting RF signals to a specific position in tissue, one can gain information related to tissue microstructure by analyzing the time series generated from the backscattered echos. In this approach, a set of six features is extracted from the spectrum of the RF time series inside a ROI. The first four features are obtained by computing the average values of the normalized spectrum in four quarters of the frequency range. The last two features are the intercept and slope of a regression line fitted to the computed spectrum. The authors have also examined the fractal dimension of the time series and discussed that these features convey tissue typing information.

4.3.4 Hybrid

The feature extraction techniques proposed in the literature take into account different properties of the ultrasound echo signals and each of them alone cannot provide a general enough description of tissue characteristics. This fact has motivated researchers to investigate the efficacy of hybrid feature vectors obtained by combining features of different nature. Fusion of diverse features extracted from Bmode or RF images as well as merging attributes extracted from different ultrasound image types have been shown to be promising methods for tissue characterization [84]. In the following we review the main investigations that have employed RF signals for creating a hybrid feature

vector.

In [99], 16 acoustical and textural parameters were extracted from signals backscattered from the prostate tissue. A stepwise parameter selection algorithm was then adopted and a subset of 10 most favorable features was used in the classification stage. Scheipers et al. [98] introduced a multi-feature algorithm for prostate diagnosis based on spectral, texture, first-order and morphological parameters. They utilized the computed features as input of two parallel neuro-fuzzy inference systems (FIS) and by combining the outputs of the FISs, made a malignancy map for highlighting cancerous regions. The incorporation of patients' clinical information into spectral feature vector was proposed in [37] for targeting and treatment monitoring of the prostate cancer. Maggio et al. [74] introduced a computer-aided detection scheme for prostate cancer diagnosis which took advantage of a combined feature vector created from spectral, statistical and textural attributes. In order to reduce the dimension of the combined feature vector and retain the most diagnostically relevant parameters, a hybrid feature selection algorithm was employed in their proposed algorithm. Recently, an ensemble learning framework has been proposed for guided prostate tissue sampling [108]. Diverse subsets of a hybrid feature vector were selected for training individual classifiers and the uncertainties concerning the labels of the data were treated by making use of a re-labeling procedure and combining the evidences raised from the individual classifiers by the evidence theory.

4.3.5 ICA Features for Medical RF Signals

After learning an ICA model (as discussed in section 1.3.2) from the set of RF signals, for computing a feature vector $f \in \mathbb{R}^N$ corresponding to a given ROI, a function $\phi(\cdot)$ should map vectors s within the ROI as,

$$f = \phi(S) \tag{4.1}$$

where $S \in \mathbb{R}^{M \times N}$ is a matrix containing M feature vectors corresponding to M RF signals inside the ROI. In this study, sum of the absolute

values of each element of s is considered as the mapping function,

$$f_j = \sum_{i=1}^{M} |s_{ij}| \tag{4.2}$$

where f_j represents the role of the jth IC in modeling signals inside the ROI. In our experiments we also examined sum of the s_{ij} values as the mapping function which did not result in favorable outcome. The inappropriateness of this mapping function can be explained by considering the possible values for the elements of s which can be both positive and negative. In this case, some of s_{ij} values can be eliminated or suppressed by the other s_{ij} quantities with opposite sign and as a consequence, the elements of the obtained vector f do not provide a proper representation of signals inside a ROI.

An important advantage of using ICA comes from the fact that it is a data-driven method. While the conventional feature extractors only use the RF echo signals inside a ROI for computing its corresponding feature vector, ICA takes advantage of the signals belong to all ROIs for the approximation of its unmixing matrix and then uses this matrix for feature computation. This characteristic of ICA provides further insight into the high generalization power of its features as compared to those of the conventional feature extractors.

4.4 Ensemble of ICA Dictionaries

4.4.1 Related Works

The idea of building an ensemble classification system using the features learned by the ICA model is relatively new. In [17], the authors addressed the small sample size problem by sampling from the original feature space. In this method several low-dimensional feature spaces are generated using the random subspace method and an ICA model is trained on each subspace. Finally, by combining the outputs of classifiers constructed on the ICA feature spaces, decisions on test samples

are made. Liu *et. al* [69] introduced an ensemble IC selection approach which employed multiple IC sets obtained by performing the ICA model with random initialization. They used a multi-objective genetic algorithm to select significant subsets from the large set of ICs and after training individual classifiers on the selected IC subspaces, aggregated their decisions to reach the final classification result. For dealing with the nongaussian process monitoring problem, it was proposed in [42] to run several ICA models using different subsets of the training data that were selected by the bagging method. Then, the ICs in each model were ranked based on their importance in monitoring and the most important ones were kept for training the individual classifiers. In a recent ICA-based process monitoring study [113], different criteria were used to select dominant ICs and the monitoring models were then built on the subspaces of the selected ICs. Outcomes of the monitoring models were then merged to achieve the final decision.

4.4.2 The Proposed Scheme

The methods reviewed in the previous section create an ensemble model by sub-sampling either from the training data or from the original or learned feature spaces. Here, we introduce a new ensemble scheme based on ICA dictionaries which takes advantage of all training data and the whole set of the learned ICs to encourage generality and preserve major information that describe the data.

Figure 4.1 represents the architecture of the proposed ensemble scheme. The main idea is to learn several distinct ICA dictionaries and construct an individual classifier using each of them. The obtained pool of classifiers is then explored to select a subset of diverse individual classifiers that complement each other and lead to an ensemble network with satisfactory performance. We used the FastICA software package for MATLAB [51] to build the ICA model. The FastICA method employs negentropy as objective function for estimating ICs and takes advantage of the fixed-point iteration algorithms for computing local extrema of $J(u)$ presented in (1.10).

In the proposed ensemble scheme, diversity among the individual

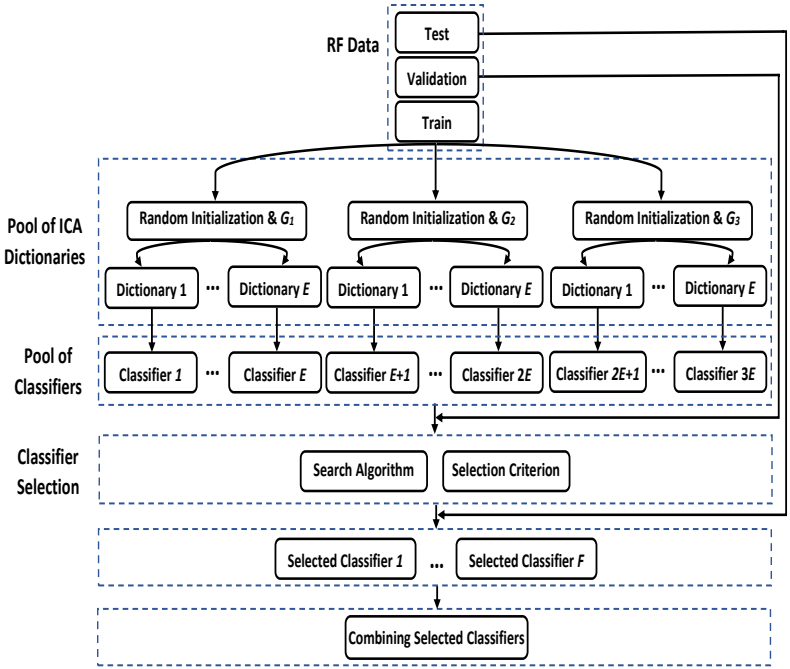


Figure 4.1: Architecture of the proposed ensemble scheme. In the train phase, $3E$ different ICA dictionaries are generated by using random initial values and three different non-quadratic functions G_1, G_2 and G_3. A classifier is built with each dictionary and a subset of $F < 3E$ classifiers is then selected in the validation phase by adopting the forward search method and a diversity-based selection criterion. In the test phase, the RF data are projected onto the spaces spanned by the dictionaries corresponding to the selected classifiers and the decisions of the classifiers are aggregated by the product rule for making the final decision.

classifiers is encouraged in two ways. First, by randomly initializing the parameters of the ICA model. Note that in addition to promoting diversity, adopting these dictionaries in an ensemble framework can significantly decrease the instability in the ICA results due to the random

initialization. The second approach is to use different non-quadratic functions G to approximate negentropy. Three non-quadratic functions have been introduced in [51] and the benefits of utilizing each of them have been reviewed. All three non-quadratic functions are used in our proposed ensemble scheme to enable it to deal with data with different characteristics. They are:

$$G_1(u_j) = \frac{1}{a_1} \log \cosh(a_1 u_j), \tag{4.3}$$

$$G_2(u_j) = -\frac{1}{a_2} \exp(-a_2 u_j^2/2), \tag{4.4}$$

$$G_3(u_j) = \frac{1}{4} u_j^2 \tag{4.5}$$

where $1 \leq a_1 \leq 2$ and $a_2 \approx 1$ are constants. It has been discussed in [51] that G_1 is a good general-purpose function, G_2 is appropriate for modeling supergaussian ICs and G_3 is proper for dealing with subgaussian ICs.

In order to choose an optimal subset of the individual classifiers from the classifier pool, a classifier selection methodology should be adopted which necessitates the incorporation of a search algorithm and a selection criterion. In this study, the forward search algorithm [46], which is known as the most intuitive greedy approach for classifier selection, is used to explore among the pool of classifiers. The advantages of using this search method has been shown in [96] based on extensive experiments and it has been demonstrated that this technique can outperform the exhaustive search algorithm which suffers from overfitting. The forward search algorithm starts with the most accurate classifier and other classifiers are added to the initial one in a sequential manner. By examining the performances of the ensemble networks made from the first classifier and any of the remaining classifiers, a classifier that leads to the best value of an evaluation criterion is selected as the second classifier. This process is repeated and finally a set of best classifiers is chosen. To assess the appropriateness of the

selected classifiers, the interrater agreement κ [65], presented in section 1.5.2, is used in this study.

After selecting a subset of the individual classifiers using the validation data, the test data are mapped onto the spaces spanned by the ICA dictionaries corresponding to the selected classifiers. Finally, decisions of the classifiers are aggregated using the sum and product rules presented in (1.27) and (1.28), respectively.

4.5 Experiments and Results

We conducted a set of experiments with clinical data as well data acquired with a tissue mimicking phantom with the following goals:

- comparing discriminatory information of the ICA features with those of the conventional feature extractors and,
- evaluating the performance of the proposed ICA-based ensemble scheme.

4.5.1 Experimental Settings

Clinical Data

The data set which has been described in section 3.4.1 was employed in our experiments. Since the pathological information is confined within the bioptic tissue sample (core), only the image portion around the biopsy needle was used to reliably associate the RF signals and the histological outcomes. Each core was then divided into 20 overlapping ROIs of size 100×11. Tumor percentages of the cores, provided by an expert physician as part of the histopathological analysis, were used for assigning class labels to them. Benign cores with tumor percentages equal to zero were considered as the healthy class and samples of the unhealthy class were selected from the cores with tumor percentages greater than 90%. For each class and in a random sampling process, 20 cores were used as training data, four cores as validation and five cores

for testing. Each of 11 RF echo signals inside a ROI was considered as a data vector. So, a set of 8800 data vectors was used for estimating ICA parameters.

The database contains several cores with tumor percentages in the interval (0,90]. This part of the data cannot be used by the classic supervised learning algorithms due to the uncertainty concerning their class labels. However, ICA can appropriately incorporate this subset of the data for computing its parameters. ICA is an unsupervised method and does not care about labels of the data. Therefore, increasing the number of data vectors helps this data-driven method to provide better estimation of the basis vectors and it brings important advantage of high generalization power to the tisseu characterization system.

The above discussion was the motivation for considering the performance of ICA with the enlarged data in our experiments where 35 uncertain cores were added to the set of training data. In this case, both groups of data with certain and uncertain class labels have been used in the procedure of estimating W. In the feature computing step, however, only samples with certain class labels were projected onto the space spanned by W.

Phantom Data

The Precision Small Parts Grey Scale Tissue Mimicking Phantom GAMMEX 404GS LE (Gammex Inc., Middleton, WI) was used. Figure 4.2 shows three grey scale targets and one anechoic target (C_4) of this phantom that were used in our experiments. They had equal sizes with a diameter of 7 mm and were located at the same depth in the phantom. The grey scale targets C_1, C_2 and C_3 had contrast values of +12, +6 and -6 dB relative to the background, respectively in which the largest value corresponded to the highest scatterer density. The anechoic target has the lowest concentration of scatterers. In our experiments, each of the four targets was considered a separate class.

The Ultrasound Advanced Open Platform (ULA-OP) [114] has been used for acquiring and recording RF signals from the targets. The central frequency of the ULA-OP system was set to 8.5 MHz and a 180-

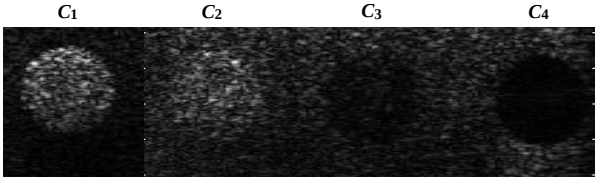


Figure 4.2: Envelope image of the phantom classes sorted based on their relative grey scales compared to the background. C_1 has the highest scatteres density while C_4 represents an anechoic target with a very low concentration of scatterers.

element linear array was used for scanning the surface of the phantom and creating 2-D RF images. A set of 70 RF images were employed in our experiments. From this set, 40, 20 and 10 images were randomly selected for training, testing and validation, respectively. This random selection was repeated five times and all results presented in the subsequent sections are the average outcomes obtained on the five sets of randomly selected data.

We investigated the performance of ICA and the conventional feature extractors by making use of three different ROI sizes. First, a rectangular window with 300 elements in the axial direction and 20 elements in the lateral direction was selected from each target. This window, which almost covered the whole target area, was then divided into non-overlapping ROIs of sizes 150×20 (large), 100×10 (medium) and 50×5 (small). In our experiments, each single ROI was considered an independent sample in the classification phase and the classifiers were trained and tested with the feature vectors computed from the utilized ROIs.

Conventional Feature Extractors

The main conventional techniques designed for extracting features from the RF echo signals were adopted in our experiments. From the family of the spectral features, the 3-dimensional vector made of Intercept, Slope and Midband fit (ISM) of the regression line fitted to the spec-

trum of the echos [72] was used. The Nakagami and GGD methods, each with two parameters related to shape and scale of the PDF, were employed from the set of the statistical features. We also used a vector composed of mean, variance and skewness of each ROI to assess the ability of lower-order statistical features and have a better understanding of the effectiveness of using ICA in exploring higher-order statistics of the RF echoes. The DWPT-based decomposition approach was selected from the group of time-frequency features. As proposed in [77], Daubechies 16 was used as mother wavelet for a 3-level decomposition of the RF signals. Then a local histogram was computed for each ROI by employing the coefficients of a 4^{th} order polynomial fitted to the results of the wavelet decomposition. Finally the standard deviations and skewnesses of the histograms were computed as features. To evaluate the performance of the multifeature approaches, another experiment was performed by concatenating all the employed conventional features into a 12-dimensional hybrid feature vector.

ICA Implementation

PCA algorithm was used in the preprocessing stage of the FastICA algorithm for computing the whitening matrix and also reducing the dimension of the data. The optimal dimension of the ICA features was determined by considering the classification results on the validation data. Based on the obtained results; the number of ICs was set to 40 for the clinical data and 46, 36 and 26 ICs were respectively selected for the large, medium and small ROIs of the phantom data.

Classifier

The DWKNN method, presented in section 1.5.1, was used as classifier in our experiments. The reason for adopting a KNN-based classifier in our experiments is that such an approach, unlike some well-known classifiers such as non-linear SVM and MLP neural network, does not map the feature vectors onto a new space for discrimination. Since the outputs of DWKNN are directly correlated with the input fea-

tures, the discriminatory information of the conventional features can be compared with the ICA features.

Ensemble ICA Model for Comparison

The ensemble model proposed by Tong *et. al* [113] was implemented and its performance was compared with that of the proposed scheme. This method uses four different criteria for selecting subsets of dominant ICs. First, a value is computed for each IC based on a given criterion and the ICs are listed in descending order according to their corresponding values. Then, l first ICs are selected for the data projection and constructing an individual classifier. The four criteria are:

- L_2 norms of the rows of W,
- L_2 norms of the columns of A,
- L_∞ norms of the learned ICs,
- nongaussianity levels of the learned ICs.

After selecting the four subsets of the ICs, the residual components are utilized to build the fifth individual classifier.

We constructed DWKNN classifiers on the subsets of ICs and combined the output weights using the sum and product rules. For adopting the best l for building the ensemble model, performance on the validation data was considered. Table 4.1 lists the examined and selected l values for the three ROI sizes.

The Proposed Ensemble Scheme

Size of the pool of ICA dictionaries was set to 30. The pool was composed of dictionaries generated by random initialization and each of the non-quadratic functions G_1, G_2 and G_3 has been used to estimate the parameters of 10 dictionaries. A pool of classifiers was then obtained by constructing DWKNNs on the generated dictionaries. In

Table 4.1: The sets of examined and selected numbers of the ICs (l) for building the individual classifiers in the Tong et. al method for the three ROI sizes

	Examined ls	Selected ls
Large ROI	[4, 6, 8, 10]	8
Medium ROI	[4, 6, 8]	8
Small ROI	[4, 6]	6

order to make a fair comparison between the proposed scheme and the Tong *et. al*'s ensemble approach, five classifiers were selected in the classifier selection phase to build the ensemble model. Note, however, that size of the ensemble can be considered as a hyperparameter of the proposed scheme and depending on the problem at hand, can be tuned using the validation data. We also utilized the 30 dictionaries incorporated in the proposed scheme to assess the performance of the classic ICA and the Tong *et. al*'s ensemble model. In this way, all the ICA-based methods were implemented with similar dictionaries.

4.5.2 Results and Discussions

Clinical Data

Figure 4.3 shows the basis vectors learned by ICA from the prostate data. Most of them are localized in time and frequency and have wavelet-like shapes. The fact that the set of basis vectors comprises diverse elements in the time and frequency axes sheds light on the capability of ICA in data modeling.

In order to have an intuitive insight into the way of modeling an echo signal with the learned basis vectors, an example is given in Figure 4.4. In this example only those basis vectors are shown that have the largest absolute weight and important contributions in the reconstruction of the RF signal. It can be seen that the basis vectors span the time axis and capture the dominant structures of the echo signal.

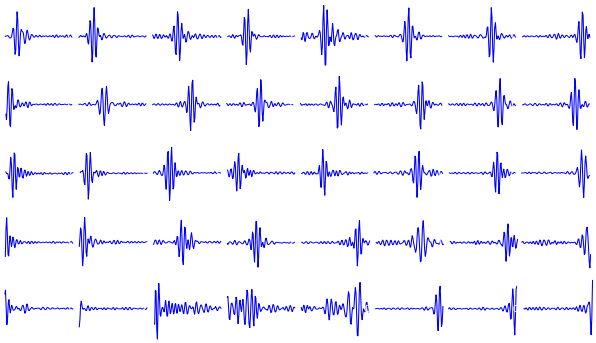


Figure 4.3: Set of the basis vectors learned by ICA from the clinical data. Most of the learned basis vectors are localized in time and frequency and have wavelet-like shapes.

Average classification results obtained with DWKNN using the different feature extraction approaches are shown in Figure 4.5. It can be seen that the augmentation of the training data by the uncertain cores leads to better performance of the ICA algorithm. It is also evident that the hybrid feature vector gives better results than the other conventional features. The performance of the hybrid vector is also comparable with that of the ICA features learned from data with certain labels but the set of ICA features obtained from the augmented data is superior to all methods.

Table 4.2 lists the best average accuracy results (%) of the studied feature extraction methods and their corresponding specificity and sensitivity values. The standard deviations of these measures are also shown. Two key issues are highlighted in this table. The first one is that by training ICA with the augmented data, the achieved results have smaller standard deviations. It again indicates that if the ICA parameters are learned by a large enough set of data examples, the resulting features yield promising classification accuracies. The second issue can be recognized by comparing the result of the hybrid feature vector with that of ICA trained on the enlarged data. Although the hybrid method gives better sensitivity than ICA, it fails to provide ac-

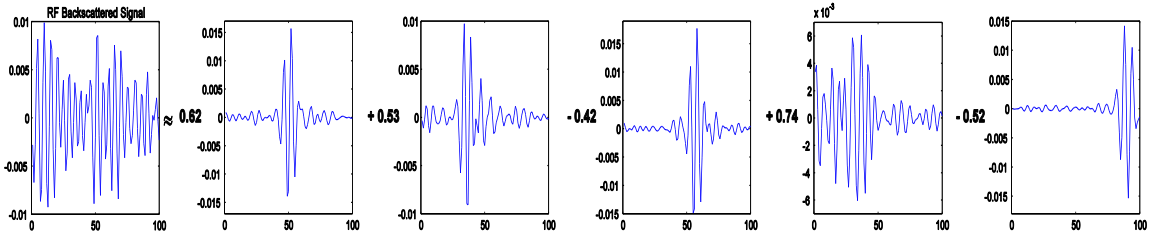


Figure 4.4: A typical RF backscattered signal and its approximation with a combination of the basis vectors which have the largest absolute weight values.

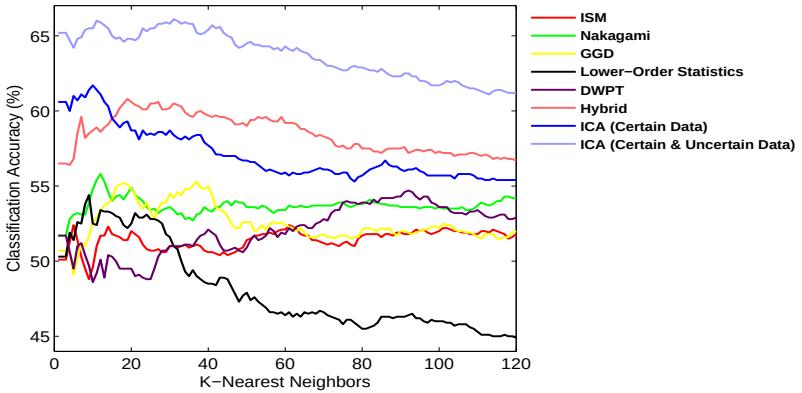


Figure 4.5: Average classification results of DWKNN on the testing prostate data using the employed feature extraction methods.

Table 4.2: Best average classification results (%) and their corresponding sensitivity and specificity values gained by the employed feature extraction techniques. Highest rates are shown in bold.

	ACC	SP	SE
ISM	52.40 ± 1.71	50.20 ± 4.82	54.60 ± 2.41
Nakagami	55.80 ± 3.49	53.80 ± 2.28	57.80 ± 7.79
GGD	55.30 ± 1.72	57.60 ± 2.61	53 ± 4.12
Lower-Order Statistics	54.40 ± 1.64	50.20 ± 3.77	58.60 ± 4.04
DWPT	54.70 ± 0.97	54.60 ± 4.62	54.80 ± 4.5
Hybrid	60.80 ± 2.28	47.40 ± 4.16	$\mathbf{74.20} \pm 2.17$
ICA (Certain Data)	61.70 ± 2.27	59.80 ± 10.64	63.60 ± 8.09
ICA (Certain & Uncertain Data)	$\mathbf{66.10} \pm 2.13$	$\mathbf{65.20} \pm 3.22$	67 ± 7.31

ceptable specificity. The ICA outcomes, however, show a satisfactory balance between specificity and sensitivity.

Phantom Data

Average classification results of DWKNN on the testing data for the three ROI sizes are plotted in Figure 4.6. Since, in general, the product rule yielded better outcomes, the results of classifier combination using this rule are presented. It is clear that by reducing the ROI size, the obtained classification results decreased. This can be explained by considering the fact that the small ROIs provide representations from the studied areas that are not as statistically significant as those of the large ROIs. Figure 4.6 shows that the results gained by the ICA features are far superior to those of the conventional feature extractors. It also highlights the advantage of ICA over the conventional feature extraction methods as far as robustness to variability in spatial resolution (ROI size) is concerned.

The results obtained by the ensemble model proposed by Tong *et. al* are compareable with those of ICA. It can be due to the fact that a small subset of the features learned by ICA is used for building the individual classifiers in this ensemble model. As mentioned before, utilizing a small portion of the features would be useful for dealing with the small sample size problem or when a few ICs can provide high discrimination power. Otherwise, reduction in the feature space size results in weak individual classifiers which do not access to enough discrimination information.

Fig. 4.6 shows that the proposed ensemble approach improved the classification performances obtained by the classic ICA features and the Tong *et. al* 's method. As mentioned earlier, all three ICA-based methods use the same set of dictionaries. However, selecting a subset of these dictionaries that provide complementary descriptions of the data and combining decisions of classifiers built with the selected dictionaries enables the proposed method to enhance the performances of the other ICA-based approaches.

For a more detailed discussion about the classification results gained by the different methods, their confusion matrices are illustrated in Table 4.3. In each matrix, the diagonal entries represent correct classification probabilities and the off-diagonal elements ($i \neq j$) are probabilities of incorrectly classifying samples of class j as class i. Note that the

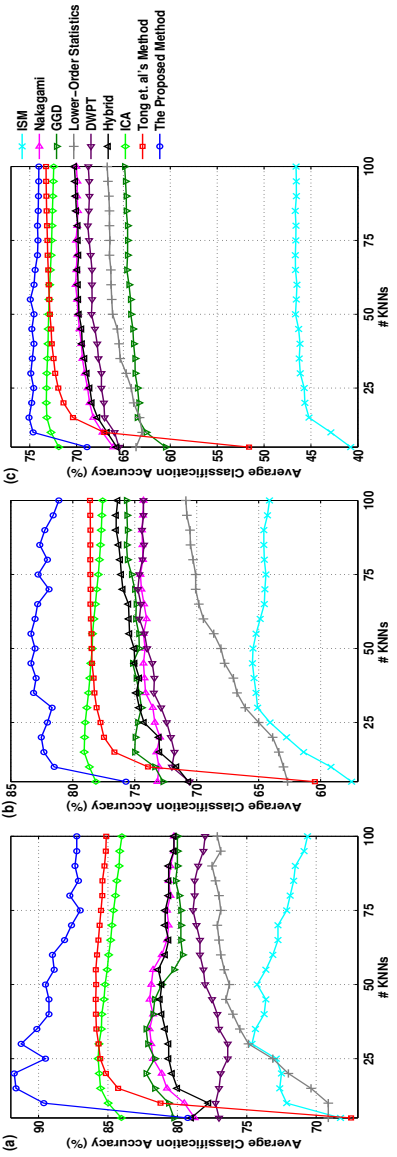


Figure 4.6: Average classification results of DWKNN on the testing phantom data using the employed feature extraction methods for (a) Large ROIs, (b) Medium ROIs and, (c) Small ROIs.

Table 4.3: Confusion matrices of the best average classification results obtained by the adopted feature extraction methods. Each column represents the classifier's performance in categorizing samples belong to one of the classes. Diagonal elements show the correct accuracy percentages while the Off-diagonal entries are miss-classification rates.

	Large ROI					**Medium ROI**					**Small ROI**				
	ACC & STD (%)	C_1	C_2	C_3	C_4	ACC & STD (%)	C_1	C_2	C_3	C_4	ACC & STD (%)	C_1	C_2	C_3	C_4
ISM	74.88 ± 0.68					65.54 ± 3.70					46.76 ± 0.79				
C_1		60.50	39	5.50	0		58.67	42.33	9.67	1.66		45.96	42.62	15.84	12.54
C_2		37	57	6	2		36.67	49	9	3.67		37.29	38.17	11.79	12
C_3		2.50	4	87.50	3.50		4.50	8	71.83	12		14.04	14.54	58.29	30.83
C_4		0	0	1	94.50		0.16	0.67	9.50	82.67		2.71	4.67	14.08	44.63
Nakagami	82.13 ± 1.85					74.67 ± 1.34					70.10 ± 1.13				
C_1		80	44.50	0	0		81.33	66.50	1	0		56.54	48.96	7.21	0
C_2		20	51.50	1.50	0		18	26.33	5.33	0		39.83	44.96	12.96	0
C_3		0	4	97	0		0.67	7.17	91	0		3.63	6.08	79.21	0.29
C_4		0	0	1.50	100		0	0	2.67	100		0	0	0.62	99.71
GGD	82.38 ± 2.98					75.75 ± 0.50					65 ± 12.48				
C_1		81	45.50	0	0		78.83	62.33	0.5	0		61.80	54.75	4.56	0
C_2		19	50	1.50	0		20.67	32	5.50	0		32.08	34.96	12.87	19
C_3		0	4.50	98.50	0		0.5	5.67	92.17	0		6.12	10.25	82.46	0.25
C_4		0	0	0	100		0	0	1.83	100		0	0.04	0.20	80.75
Lower-Order Statistics	77.50 ± 3.95					71.12 ± 3.81					66.80 ± 2.74				
C_1		90	64.50	0	0		80.33	70	0.33	0		65.25	59.63	6.5	0
C_2		10	30	0.5	0		18.67	21.33	4.83	0		29.63	31.12	10.71	0
C_3		0	5.50	90	0		1	8.67	82.84	0		5.12	9.25	71.12	0.33
C_4		0	0	9.50	100		0	0	12	100		0	0	11.67	99.67
DWPT	78.87 ± 0.52					74.79 ± 0.95					68.81 ± 0.84				
C_1		88.50	63	0	0		73.17	56.66	1	0		67.50	60.58	9.67	0
C_2		11.50	32.50	5.50	0		25.33	38.17	11	0		27.50	30.92	13	0
C_3		0	4.50	94.50	0		1.50	5.17	87.83	0		5	8.50	77.12	0.29
C_4		0	0	0	100		0	0	0.17	100		0	0	0.21	99.71
Hybrid	81.50 ± 1.80					76.58 ± 1.05					70.29 ± 0.67				
C_1		85.50	54	0	0		84.83	65.83	0.34	0		58.12	50.12	6	0
C_2		14.50	42.50	1.50	0		14.17	27.67	4.83	0		37.79	42.79	12.66	0
C_3		0	3.50	98	0		1	6.50	93.83	0		4.09	7.09	80.42	0.17
C_4		0	0	0.5	100		0	0	1	100		0	0	0.92	99.83
ICA	85.74 ± 2.27					79.13 ± 1.64					73.23 ± 1.08				
C_1		70.48	24.27	0	0		60.89	34.59	0.11	0		50.35	37	1.14	0
C_2		29.52	72.63	0.15	0		37.71	56.91	0.57	0		41.57	47.93	2.31	0
C_3		0	3.10	99.85	0		1.41	8.49	98.75	0.02		8.08	15.07	94.99	0.33
C_4		0	0	0	100		0	0	0.56	99.98		0	0	1.56	99.67
Tong *et. al*'s Method	85.93 ± 3.05					78.62 ± 1.74					73.28 ± 1.10				
C_1		73.43	27.60	0	0		54.86	33.12	0.19	0		49.78	38.96	2.46	0.01
C_2		26.57	70.85	0.55	0		44.82	62.10	1.72	0		46.49	52.79	5.20	0
C_3		0	1.55	99.45	0		0.32	4.77	97.55	0.03		3.73	8.24	91.21	0.61
C_4		0	0	0	100		0	0	0.54	99.97		0	0	1.12	99.37
The Proposed Method	91.75 ± 2.10					83.37 ± 1.29					75.10 ± 2				
C_1		83.50	13.50	0	0		67.83	25.50	0	0		53.37	33.62	0.66	0
C_2		16.50	83.50	0	0		30.83	66.17	0.17	0		39.25	51.33	1.87	0
C_3		0	3	100	0		1.33	8.33	99.50	0		7.35	15.05	96.08	0.37
C_4		0	0	0	100		0	0	0.33	100		0	0	1.37	99.62

matrices represent the best average classification rates of each method. The confusion matrices show that categorizing samples belonging to C_1 and C_2 is difficult for all methods and these classes caused the highest misclassification rates. This result is not surprising because these two classes have close scatterers densities and the features extracted from them form a highly overlapping area in the feature space. Table 4.3 shows that classifying samples of C_3 is not as difficult as the first two classes. It can be described by the reduction in the scatterers concentration in C_3 which makes it more distinct from C_1 and C_2. As can be seen from Figure 4.2, the contrast of C_4 is much more smaller than the rest of the classes and it has the lowest scatterers density. So, the overlap of the C_4's features with those of the other classes is negligible and they can be easily categorized by the classifiers. By considering the classification results gained by the conventional features for the first two classes, it can be noticed that the classifiers could not provide balanced rates which is due to significant overlap of the features. Although it is not the case for ISM, it failed to yield high classification accuracies. By using the ICA features, however, a good compromise between balance and accuracy of the classification rates was achieved that indicates the learned features provide considerable discriminatory information and a good separation between the two classes.

There is a direct correlation between the success of the proposed scheme and the optimality of the features provided by the selected ICA dictionaries. To study the advantage of utilizing the dictionary/classifier selection strategy in the proposed ensemble model, the effect of adding each of the five selected dictionaries to the ensemble model is illustrated in Figure 4.7. In this figure, the curves of Classifier 1 belong to the first selected dictionaries and the rest of the curves show the trend in the classification results obtained by increasing the ensemble models' sizes. It can be seen that for all the three ROI sizes, the inclusion of every new selected dictionary in the ensemble model enhances the classification results gained in the previous stage. This means that the ICA dictionaries complement each other and implies that the selection procedure, composed of the forward search algorithm and the diversity-based criterion, worked well.

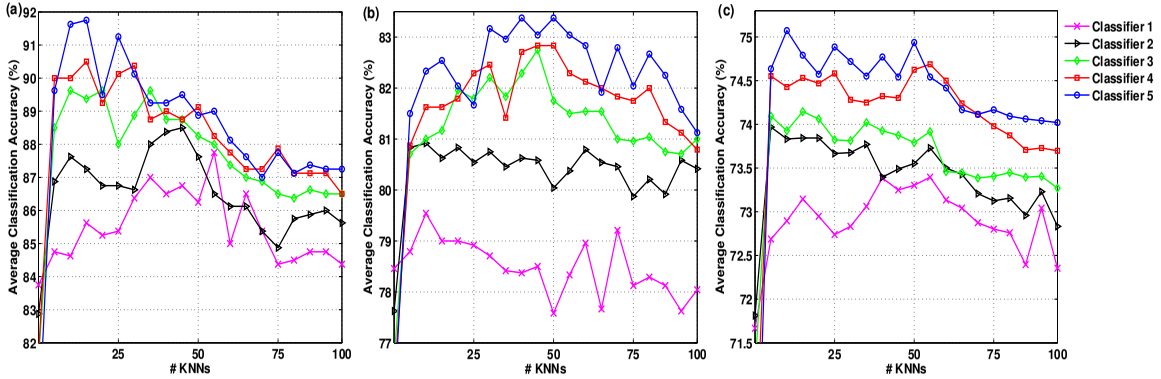


Figure 4.7: Effects of adding the selected classifiers/dictionaries to the structure of the proposed ensemble model where the curve of Classifier 1 shows the outcomes obtained by the first selected classifier, the curve of Classifier 2 represents the performance of the ensemble model with the first and second selected classifiers and so on. (a) Large ROIs, (b) Medium ROIs and, (c) Small ROIs.

4.6 Conclusion

The issue of classifying medical RF ultrasonic signals was addressed in this study. It was proposed that the learning strategies are viable alternatives to the conventional methods designed for extracting feature from the medical RF signals and the ICA model was employed for learning features from the RF echoes. Several different ICA dictionaries were then trained with the RF data and a subset of them that provided complementary descriptions of the backscattered signals were incorporated into the structure of an ensemble classification model. The classic ICA method and the proposed ensemble model were respectively evaluated with a clinical and a tissue mimicking phantom data. It has been shown that the features learned by ICA can yield better classification results that their conventional counterparts and a combination of ICA dictionaries that complement each other would provide more satisfactory outcomes than a single ICA dictionary.

Part II

Statistical Modeling of Left Ventricular Mechanics and Shape to Detect Myocardial Infarction

Chapter 5

Echocardiographic Deformation Imaging

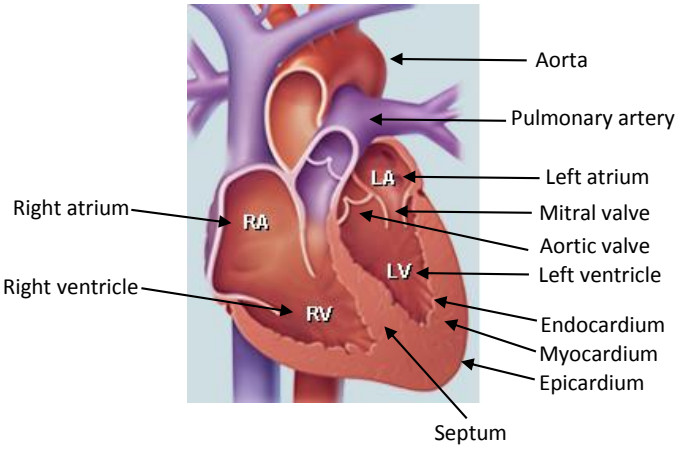


Figure 5.1: Heart cross section.

5.1 Cardiac Anatomy

The human heart (Figure 5.1) weights around 300 grams and it typically has the size of a closed fist. It is surrounded by a double-layered sac called the *pericardium*. The heart wall is divided into three layers. The outermost layer is called *epicardium* which protects the heart. The middle layer is composed of cardiac muscle cells and it is called *myocardium*. This layer is responsible for pumping the blood throughout the body. The *endocardium* is the innermost layer of the heart which directly contacts the blood and covers the heart valves.

As shown in the cross section representation of the heart in Figure 5.1, it is composed of four chambers including two upper chambers called the *right atrium* (RA) and *left atrium* (LA) and two lower chambers called the *right ventricle* (RV) and *left ventricle* (LV). The left and right sides of the heart are separated by a wall known as *septum*.

The right side of the heart collects the deoxygenated blood from the body and pumps it to the lungs through the *pulmonary artery* (Figure 5.1) and then the left heart pumps the oxygenated blood into

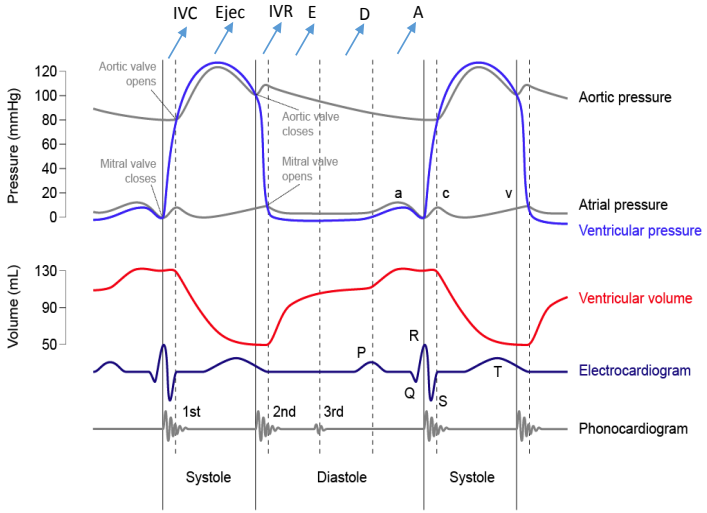


Figure 5.2: Schematic Wiggers diagram showing the cardiac cycle together with its six mechanical phases and the ECG [92]. Timings of opening and closing of aortic and mitral valves along with aortic, atrial and ventricular pressures and ventricular volume are also depicted

the *aorta*.

5.2 The Cardiac Cycle

The cardiac cycle can be divided into two major time phases namely *systole* and *diastole*. Diastole, which usually takes more than 50% of the cardiac cycle (Figure 5.2), includes all the events associated with filling of the ventricles while systole includes all the events associated with ejecting blood from the ventricles.

The period between the onset of the QRS complex on the electrocardiography (ECG) and mitral valve closure is referred to as the *electromechanical coupling* (EMC) phase. Following this phase, both

left ventricular valves are closed and the corresponding time period is known as *isovolumetric contraction* (IVC). The next phase is *ejection* (Ejec) when the aortic valve is open and blood rapidly ejects from the ventricle into the arteries. This phase ends by aortic valve closure. The diastolic period then begins with the *isovolumetric relaxation* (IVR) phase during which ventricular pressure drops below the LA pressure resulting the mitral valve opening. After this phase, blood rapidly flows into the ventricle which is referred to as *early filling* (E) or rapid inflow. In the next phase of diastole, filling of the ventricle occurs more slowly which is known as *diastasis* (D). In the last phase of diastole which is called *atrial/late filling* (A), slow filling of the ventricle happens due to atrial contraction. Note that, the onset of atrial contraction can be defined by the onset of the P-wave on the ECG.

5.3 Myocardial Deformation Estimation

Ultrasonic deformation estimation is a noninvasive technique for the assessment of regional myocardial function [29, 106]. This (relatively) new technique, is based on *strain* and *strain rate* measurements as concepts derived from mechanical engineering. Strain is a dimensionless parameter which is defined as the deformation of an object normalized to its original shape. In the case of a one-dimensional object, in which the only possible deformation of the object is lengthening or shortening (Figure 5.3), strain can be written as,

$$\varepsilon = \frac{L_2 - L_1}{L_1} = \frac{(y_2 - x_2) - (y_1 - x_1)}{y_1 - x_1} \tag{5.1}$$

where ε is the strain symbol and (x_1, y_1) and (x_2, y_2) represent the position of a specific point in the object before and after deformation, respectively.

Strain rate can be defined as the speed in which the deformation (i.e strain) occurs and is computed by dividing both sides of equation

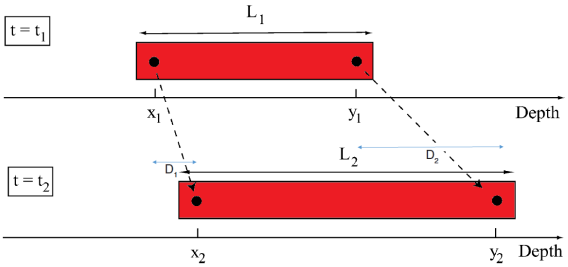


Figure 5.3: Schematic illustration of the way in which strain and strain rate measures are computed. An object with the original length of L_1 is deformed from time t_1 to t_2 to have length L_2. Strain and strain rate are defined by considering the change in the length of the object and the time duration that this lengthening/shortening is happened.

(5.1) by $\Delta t = t_2 - t_1$ as the time duration of the deformation,

$$SR \approx \frac{\varepsilon}{\Delta t} = \frac{\frac{(y_2-y_1)}{\Delta t} - \frac{(x_2-x_1)}{\Delta t}}{y_1 - x_1} = \frac{\frac{D_2}{\Delta t} - \frac{D_1}{\Delta t}}{y_1 - x_1} \approx \frac{v_2 - v_1}{L_1}. \tag{5.2}$$

To describe deformation patterns of the heart as a 3-dimensional object, a local heart coordinate system has been introduced. Figure 5.4 represents this coordinate system which includes three mutually perpendicular axes known as the *radial* (R), *longitudinal* (L) and *circumferential* (C). These three axes make the bases of normal strain (rate) measurements.

5.4 Myocardial Deformation Imaging

Echocardiographic strain (rate) traces are measured in clinical equipment by using *tissue Doppler imaging* (TDI) [30, 47] or *speckle tracking imaging* (STI) [90] techniques.

In the TDI technique, regional instantaneous myocardial velocities can be measured by colour Doppler. Hereto, by integrating the regional

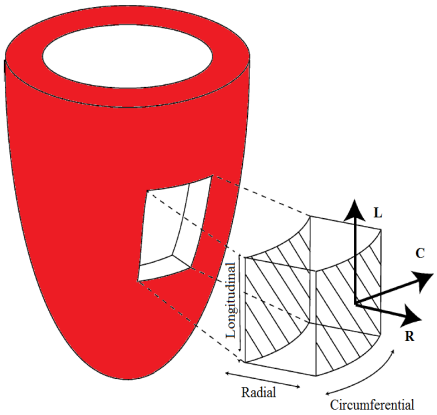


Figure 5.4: Representation of the local heart coordinate system consisting of radial, longitudinal and circumferential axes.

velocity profiles, their corresponding strain rate curves can be achieved. The main advantage of TDI is its high frame rate (> 180 Hz) which allows to assess short-lived events during the cardiac cycle. It also provides a fast qualitative analysis as deformation traces can be easily derived by moving the mouse pointer over the myocardium and as a result, no postprocessing is required [73]. This technique, however, can only measure myocardial deformation in the direction of the ultrasound image line. As a consequence, the obtained deformation traces are dependent on the angle between the ultrasound wave and the target segment [30, 59]. Since the strain estimation in this technique depends on the ultrasound line's direction, and since there are limited number of views for visualizing the heart, only L strain component can be measured in all LV segments.

The STI technique, which is a more recent approach to the deformation analysis, is based on tracking a unique speckle patterns (i.e. a ROI) on the myocardial wall from one B-mode image frame to the next one (Figure 5.5). Having the ROI's position information in two successive frames at hand, its relative displacement per unit distance is computed. This is equivalent to instantaneous strain of a tissue

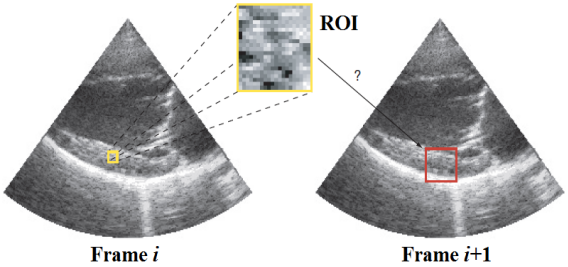


Figure 5.5: An example illustrating the principle of speckle tracking. A ROI on the cardiac wall is tracked from frame i to frame $i+1$.

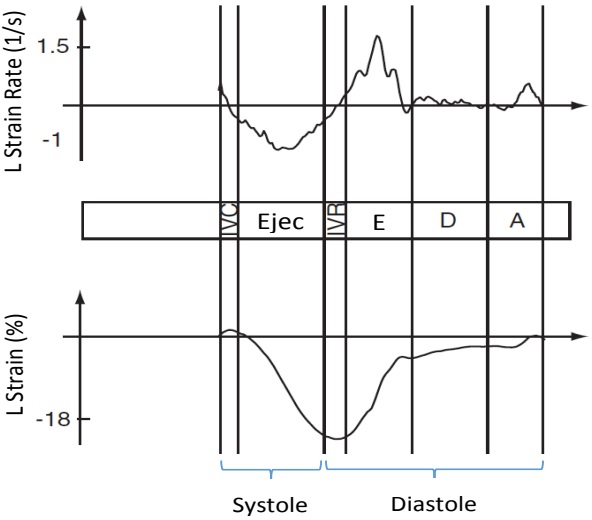


Figure 5.6: Examples of L strain and strain rate curves measured in the LV along with the timing of cardiac mechanical events.

segment. To have the strain curve corresponding to the whole cardiac cycle, the instantaneous strain measures should be accumulated. A strain rate curve can then be obtained by computing the temporal derivative of this strain curve. Tracking speckle in the B-mode image brings the advantage of angle independency to the strain (rate) estimation system. As a result, all strain components (i.e. R, L and C) can be measured with this technique. The major disadvantage of STI, however, is its need to images with high spatial resolution (i.e high line density) for efficient speckle tracking. This requirement limits the frame rates to 50-70 Hz and therefore results in relatively low temporal resolution.

Figure 5.6 illustrates examples of measured L strain and strain rate curves along with timing of the cardiac mechanical events.

Chapter 6

Principal Component Analysis for the Classification of Cardiac Motion Abnormalities Based on Echocardiographic Strain and Strain Rate Imaging

This chapter is based on a paper published in *The 8th Int. Conf. on Functional Imaging and Modeling of the Heart (FIMH), 2015*, M. Tabassian, M. Alessandrini, L. De Marchi, G. Masetti, N. Cauwenberghs, T. Kouznetsova, and J. D'hooge, "Principal Component Analysis for the Classification of Cardiac Motion Abnormalities Based on Echocardiographic Strain and Strain Rate Imaging", pp. 83–90.

6.1 Abstract

Clinical value of the quantitative assessment of regional myocardial function through segmental strain and strain rate has already been demonstrated. Traditional methods for diagnosing heart diseases are based on values extracted at specific time points during the cardiac cycle, known as 'techno-markers', and as a consequence they may fail to provide an appropriate description of the strain (rate) characteristics. This study concerns the statistical analysis of the whole cardiac cycle by the PCA method and modeling the major patterns of the strain (rate) curves. Experimental outcomes show that the PCA features can outperform their traditional counterparts in categorizing healthy and infarcted myocardial segments and are able to drive considerable benefit to a classification system by properly modeling the complex structure of the strain rate traces.

6.2 Introduction

Echocardiography is the modality of choice in clinical diagnostics and for the noninvasive assessment of heart function. In daily clinical practice, visual evaluation is widely used to determine regional abnormalities in myocardial wall motion. Although this qualitative assessment can be done easily, it suffers from significant inter-observer variability which reduces its clinical value.

Tissue Doppler imaging and speckle tracking are two promising echocardiographic techniques that have been developed for the noninvasive study of myocardial function. Based on these techniques, strain (rate) imaging has been introduced to provide an effective approach for the assessment of changes in the regional myocardial wall motion and deformation [29]. End-systolic strain and peak-systolic strain rate are two traditional techno-markers that have been extensively used by clinicians to describe the strain (rate) profiles and to classify different heart diseases [48]. These traditional features, however, ignore the diastolic period of the cardiac cycle. They also represent the value of

the strain (rate) profile at only one time point and as a result, cannot capture the temporal information available in the deformation curves.

Despite several studies for the classification of regional myocardial function based on the traditional features of the strain (rate) curves [48, 57], only a few investigations have been carried out for the detection of heart abnormalities by taking into account the whole temporal behavior of the strain (rate) curves. The idea of analyzing the systolic phase of strain curves, derived from tagged magnetic resonance imaging, by PCA was initially proposed in [19]. The authors showed that the statistical reference model achieved by employing PCA and normal strain curves can properly detect abnormal strain patterns. Inspired by [19], PCA has been used in [2] to model ultrasonic strain and strain rate traces of the healthy subjects and it has been discussed that the PCA attributes can provide more information about strain (rate) curves than the traditional features. In [79], an artificial neural network (ANN) was used to classify strain profiles obtained at baseline and during experimentally induced acute ischemia using animal data. In a pre-processing step, each strain curve was represented by 70 equidistant samples and normalized in amplitude. The obtained profiles were then given to an ANN for categorization.

Following the results presented in [2], this paper addresses the subsequent issues: 1) building two PCA models by making use of the normal and acute infarcted strain and strain rate curves, 2) incorporating the PCA and traditional features in a classification system and examining their capabilities for the categorization of normal and acute infarcted strain and strain rate traces and, 3) comparing the PCA features extracted from the strain and strain rate traces in terms of the amount of the discriminatory information that they provide for a classification system.

6.3 Materials and Methods

6.3.1 Data Acquisition and Preprocessing

A group of 27 normal subjects and 54 subjects with acute myocardial infarction was used in this study. For the patients, myocardial segments were categorized into infarct, border and remote based on MRI-delayed enhancement and the perfusion territory of the infarct-related vessel [48]. Data acquisition was performed at high frame rate (>180 Hz) with a GE VingMed Vivid7 equipped with a 2.5 MHz transducer. For each subject, data were acquired in the apical 2-, 3- and 4-chamber views with optimization of the pulse repetition frequency in order to avoid aliasing. An event-driven graphical user interface called SPEQLE [20] was used for the post-processing of the data to extract longitudinal strain (rate) traces in an 18-segment model of the left ventricle [16]. Since the number of samples of the extracted curves could be different due to the differences in the heart rates of the subjects, a linear interpolation procedure was adopted to have the same number of samples in all traces. To avoid unwanted changes of the curves due to the interpolation procedure, each of the six cardiac phases (i.e. electromechanical coupling, isovolumetric contraction, ejection, isovolumetric relaxation, early filling and late filling [29]) was interpolated separately and then merged to have the whole heart cycle. The interpolated curves were then used in the PCA implementation and classification phases.

Table 6.1 lists the number of subjects that were selected randomly from the healthy and pathological groups for building training, validation and test sets. This random selection was repeated 10 times and the results presented in section 6.4 are the average of running the classifier on these 10 different sets of data. Since for a pathological subject only the subset of acutely infarcted segments was used for the classification task, the number of utilized pathological subjects in Table 6.1 was more than the healthy ones so that both groups had roughly the same number of curves in the training, validation and test sets. Note that, the segmental strain and strain rate curves were sorted in two different groups of training, validation and test data to study their clinical relevance for discriminating normal and infarcted traces.

Table 6.1: Number of subjects taken from the healthy and pathological groups for the training, validation and test sets.

	#Training	#Validation	#Test
Healthy Subjects	12	5	10
Pathological Subjects	25	9	20

6.3.2 Classifier

Performance of a classification system depends on both the features extracted from the data and the classification technique. In order to differentiate between the effects of the employed features and the classification strategy on the final classification outcomes, two different classifiers namely LMB [82] and SVM [25, 117] are used in our experiments.

6.4 Classification Results

Average classification results on the test data for LMB and SVM are shown in Figure 6.1 and Table 6.3, respectively. Table 6.2 gives the best average classification results and their corresponding sensitivity and specificity values obtained by the LMB and KNNs from [1 100] interval. The number of PCs was set to 10 based on the favorable results gained with the validation data.

6.5 Discussion

Tables 6.2 and 6.3 show that regardless of the classification methodology, the PCA features yielded more satisfactory accuracy rates for both the strain and strain rate traces compared to the traditional features. This means that by exploring the whole cardiac cycle instead of only

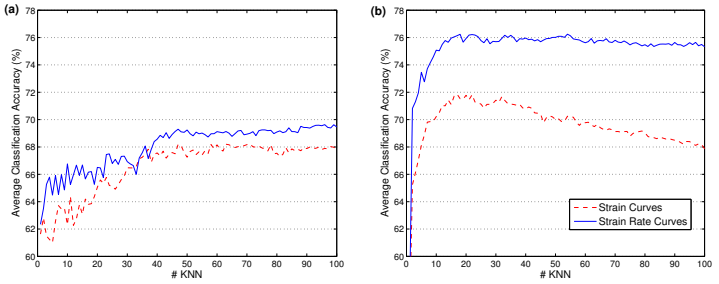


Figure 6.1: Average classification results (%) on the test strain and strain rate curves with the LMB method using (a) the traditional features and, (b) the PCA features.

Table 6.2: Best average classification results (%) and their corresponding sensitivity and specificity values of the LMB method with the traditional and PCA features on the test strain and strain rate curves.

	ACC	SE	SP
Traditional Features			
Strain	68.22	67.51	68.86
Strain Rate	69.63	73.35	66.35
PCA Features			
Strain	71.86	64.98	78.90
Strain Rate	76.26	69.93	83.44

Table 6.3: Average classification results (%) and their corresponding sensitivity and specificity values of the SVM method with the traditional and PCA features on the test strain and strain rate curves.

	ACC	SE	SP
Traditional Features			
Strain	69.61	55.14	83.39
Strain Rate	71.17	67.23	75.56
PCA Features			
Strain	70.06	73.41	67.07
Strain Rate	74.32	77.03	72.28

the systolic period, considerably more information about the characteristics of the strain and strain rate curves can be obtained. Another key point that is highlighted in Tables 6.2 and 6.3 is the difference between the classification accuracies obtained with features extracted from the strain and strain rate traces. While by using the PCA features, the classification accuracies on the strain rate curves are remarkably better than those of the strain traces, there is not much benefit to be gained by employing the traditional features and the accuracies obtained for the strain rate traces are slightly better than the strain curves. This observation suggests that the strain rate traces provide more discriminatory information than the strain curves and these extra information could be captured by an efficient statistical model like PCA.

Tables 6.2 and 6.3 also show that for both the PCA and traditional features, the sensitivity and specificity rates achieved by LMB and SVM varied considerably. It means that the classification strategy has a direct effect on the final outcomes and suggests that different classification methodologies should be examined to find a suitable setup. These results also imply that by combining several classifiers that address the classification problem from different points of view, a good compromise between the sensitivity and specificity values could be obtained that is a topic for the future research.

6.6 Conclusion

In this study, it was hypothesized that the temporal behavior of the segmental strain (rate) curves contains valuable diagnostic information which can be captured by a rigorous statistical approach. The PCA method was then employed to statistically analyze the strain (rate) traces. In order to evaluate the usefulness of the PCA features, they were compared with the end-systolic strain and peak-systolic strain rate values as the traditional features. Experiments with a data set of strain (rate) curves of healthy and pathological subjects demonstrated that the PCs can provide more discriminatory information for the classification system than the traditional features. Our experiments also showed that analyzing the strain rate traces with PCA would lead to better results than the strain curves. However, more thorough analysis with a larger set of data is needed to improve the classification performance and to determine the role of different parameters that affect the obtained outcomes.

Chapter 7

Automatic Detection of Myocardial Infarction by Spatio-Temporal Analysis of Echocardiographic Strain and Strain Rate Curves

This chapter is based on: 1) a paper published in *IEEE Int. Ultrason. Symp. 2015*, M. Tabassian, M. Alessandrini, L. Herbots, O. Mirea, J. Engvall, L. De Marchi, G. Masetti, and J. D'hooge, "Automatic detection of ischemic myocardium by spatio-temporal analysis of echocardiographic strain and strain rate curves" and, 2) a paper which is under preparation for submission to a journal.

7.1 Abstract

Interpretation of ultrasonic deformation traces for making a diagnosis on local myocardial function has been known to be a challenging task in daily clinical practice. A traditional approach is to use values extracted at specific time points during the cardiac cycle which has the main drawback of not taking the temporal information of the deformation traces into account. This study presents a framework for the automatic detection of ischemic myocardium by statistical analysis of the entire segmental strain and strain rate curves using PCA. Having the PCA-derived parameters of the regional temporal profiles at hand, a spatio-temporal representation of the global left ventricle (LV) function is established to train a classification system. Experimental outcomes show that the proposed deformation representation of the LV can outperform its traditional counterpart in categorizing healthy from ischemic myocardium.

7.2 Introduction

Echocardiographic strain and strain rate imaging has been known to be a well-suited noninvasive method to the diagnosis of heart diseases that result in regional dysfunction of the ventricle. Interpretation of strain (rate) curves, however, is a major challenge in the diagnostic procedure and its clinical value depends on the cardiologist's experience. With the aim of describing the strain (rate) profiles and classifying different heart diseases, some traditional techno-markers (e.g. endsystolic strain, peak-systolic strain (rate)) have been used extensively [48]. These traditional techno-markers, however, ignore the diastolic period of the cardiac cycle and represent the value of the strain (rate) profile at only one time point. As a result, they cannot capture the temporal information available in the deformation curves.

In this context, Clarysse et al. [19] proposed to model the systolic phase of strain curves, extracted from tagged magnetic resonance data, by a statistical approach instead of utilizing the traditional techno-markers. Inspired by their work, the whole temporal profiles of the

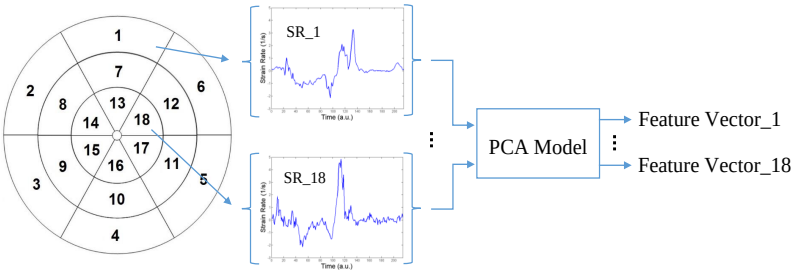
echocardiographic strain (rate) measurements of the left ventricle (LV) were modeled by PCA [2]. The idea was to build a statistical model of the normal deformation patterns from healthy subjects and use the parameters of this model to detect abnormalities in patients' deformation curves. Abnormality detection was performed by projecting a given curve onto the subspace spanned by the PCA parameters and then building a map (i.e. Bull's eye) that reveals distances between the parameters of the segments and the normal model. This distance map was then presented to a cardiologist in order to be classified as normal or pathologic. Although this work takes advantage of the entire deformation profiles for modeling their temporal patterns, it lacks an objective and automatic scheme for disease classification. This was the motivation for developing a framework in [110] for the automatic detection of abnormalities in the segmental deformation traces. After building a PCA model with the deformation profiles of healthy and pathological subjects, a classifier was trained with the extracted temporal parameters. The obtained results confirmed the utility of the PCA-based classification system and showed that it could yield better results than the same classifier trained with the traditional technomarkers.

The current study moves the framework presented in [110] one step further by proposing a global representation of the LV function. This is done by concatenating the PCA-derived temporal parameters of the LV segments to model their mechanical interactions.

7.3 Materials and Methods

7.3.1 Data Acquisition and Preprocessing

A set of 60 normal subjects from the DOPPLER-CIP multicenter clinical study (`http://www.dopplercip.be`) and 60 patients with acute infarction from the stem cell [58] and salvage [28] studies, which have been conducted in the University Hospital Leuven, was used in this study. Data acquisition and preprocessing is the same as described in section 6.3.1.



(a)

{Feature Vector_1, …, Feature Vector_18}: Spatio-Temporal Feature Vector

(b)

Figure 7.1: Schematic illustration of the first approach proposed in this study for modeling global LV function. (a) Two independent PCA models are constructed with the strain and strain rate curves of the LV segments to capture their major temporal patterns. For each segment, its strain and strain rate curves are then projected onto the subspaces spanned by their corresponding set of PCs to have a vector of features that represents the function of that segment. (b) Feature vectors of the LV segments are concatenated into a single spatio-temporal feature vector for building a classification system.

7.3.2 Spatio-Temporal Deformation Representation

In this study, two different PCA-based methodologies for modeling spatio-temporal (S-T) parameters of the LV function are proposed.

Methodology I

Regional deformation profiles of both normal and ischemic groups were used to build two separate PCA models for the strain and strain rate traces. A subset of the M first PCs with the largest eigenvalues were retained from each model to ensure that most of the temporal variations of the data were captured. Then, every strain (rate) curve x_n was

projected onto the subspace spanned by its corresponding set of PCs to have the coefficient set $A_n = \{a_{n1}, \ldots, a_{nM}\}$ (Figure 7.1(a)). Since the obtained vectors of coefficients are different for different traces, they can be used as the feature vectors that represent the temporal behavior of the deformation traces.

For each subject, temporal feature vectors of its LV segments, according to 18-segment model [16], were concatenated into a single feature set $\{A_1, \ldots, A_{18}\}$ (Figure 7.1(b)) where A_n is the set of temporal coefficients belong to the nth segment. This new feature set accounts for the mechanical interactions between the LV segments thereby providing a spatio-temporal representation of the global LV deformation.

Methodology II

In the second approach, 18 curves of each subject were first concatenated to form a single vector $CRV = \{curve_1, \ldots, curve_{18}\}$. A PCA model was then built using the CRV vectors obtained from all subjects (Figure 7.2). Since the CRV vectors contain implicitly the temporal as well as spatial information of the deformation traces, the PCA model can give a S-T representation of LV function. Note that, like the first methodology, two independent PCA models were built for the strain and strain rate curves.

7.3.3 Feature Selection

The P-metric method [55], presented in section 1.4, was used to select relevant PCs of each approach that provide significant discrimination information.

As mentioned in section 6.3.1, LV segments of the patients were categorized into infarct, border and remote based on MRI-delayed enhancement information. In methodology I, only the strain (rate) curves corresponding to the infarcted LV segments were taken from the patient class to perform feature selection. The reason for doing so was to have a proper representation of the abnormal class and to select those PCs that can provide significant discriminatory information. In

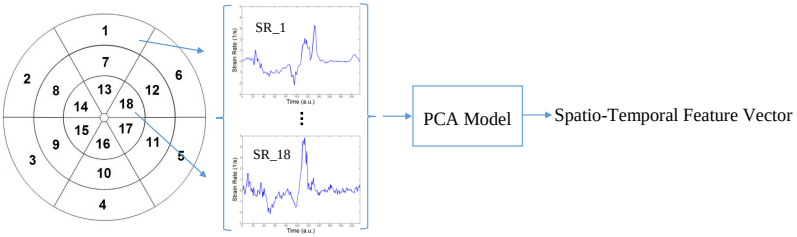


Figure 7.2: Schematic illustration of the second proposed approach for modeling global LV function. For each subject, its 18 strain (rate) curves are concatenated to form a new vector that consist temporal as well as spatial information of the deformation profiles. Two independent PCA models are then constructed with the new vectors built from the connected strain and strain rate curves.

methodology II, however, the CRV vectors include traces of the infarcted segments so the labels of the subjects can be used in the feature selection procedure.

7.3.4 Parameter Settings with Cross-Validation

The cross-validation (CV) technique was adopted to set the optimal number of first or selected PCs that should be retained and to ensure that the classification results are not biased towards a subset of the subjects. Experiments were carried out with 10-fold CV. Hereto, subjects were randomly divided into 10 equal-size folds. Nine folds were used for training and the last one was used for testing the classifier.

7.3.5 Traditional Features

To assess the benefit of the spatio-temporal features derived from the PCA method as compared to the traditional markers, two different sets of features were generated by concatenating:

1. End-systolic strain and,

2. Peak-systolic strain rate values

of the segments belong to each LV.

7.3.6 Classification

The KNN method with the *cosine distance metric* was used to categorize the S-T and traditional features of the normal and ischemic subjects.

7.3.7 Cardiologist Reading

Visual assessment of wall motion and wall thickening based on grayscale ultrasound images is currently the most common clinical routine for the evaluation of myocardial function [29, 106]. Another common method for the assessment of the regional heart function is to use a bull's eye representation of end-systolic strain values [48].

To contrast the performance of our proposed automatic approach with the outcome of a typical visual assessment procedure, we asked a cardiologist to make a decision on the LV function (i.e. normal or infarcted) of a given subject based on:

- Six grayscale wall motion videos acquired from: anterior and inferior apical 2-chamber, anteroseptal and inferolateral apical 3-chamber and anterolateral and inferoseptal apical 4-chamber views (Figure 7.3)
- A bull's eye plot of the end-systolic strain values (Figure 7.4)
- The acquired strain and strain rate curves of 18 segments of the LV (Figure 7.5)

A subset of 20 normal subjects and 20 subject with myocardial infarction (MI) was randomly selected from the original set of data for the cardiologist reading. The items listed above were presented to

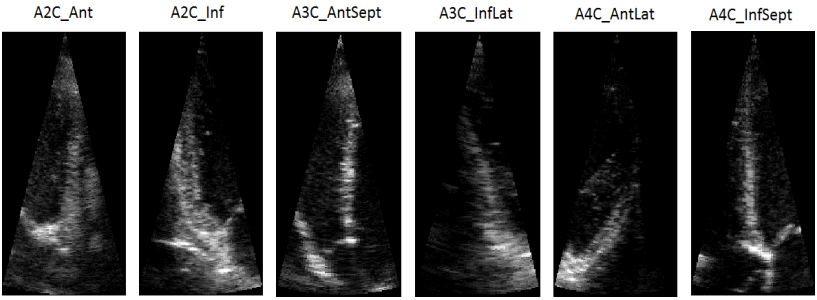


Figure 7.3: Six garyscale wall motion videos acquired from anterior and inferior apical 2-chamber, anteroseptal and inferolateral apical 3-chamber and anterolateral and inferoseptal apical 4-chamber views. The videos were presented to the cardiologist for decision making.

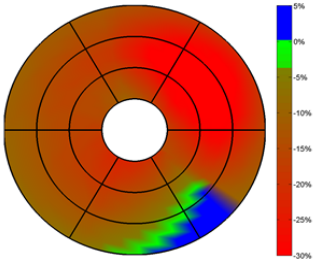


Figure 7.4: Example of a bull's eye plot generated by the end-systolic strain values of the LV segments for cardiologist reading.

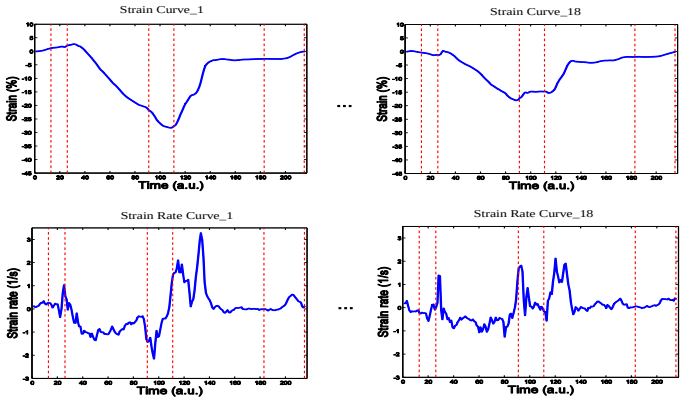


Figure 7.5: Samples of the strain and strain rate curves of one of the subjects prepared for cardiologist reading.

the cardiologist in two separate phases. In the first phase, only the wall motion videos were demonstrated to the cardiologist for making a decision while in the second phase, the cardiologist could use all listed items for decision making.

7.4 Results and Discussion

7.4.1 PCA Results

The first three modes of variation of the strain and strain rate PCs of the first PCA model are illustrated in Figure 7.6. In order to investigate the PCs' structures in terms of timing of the six mechanical phases of the cardiac cycle, the timing of the onset of each phase is shown with a red vertical line.

For the strain curves (Figure 7.6(a)), the first PC explains the amplitude variations among the subjects. The second PC describes variations in timing of shortening and lengthening of the segments. It shows

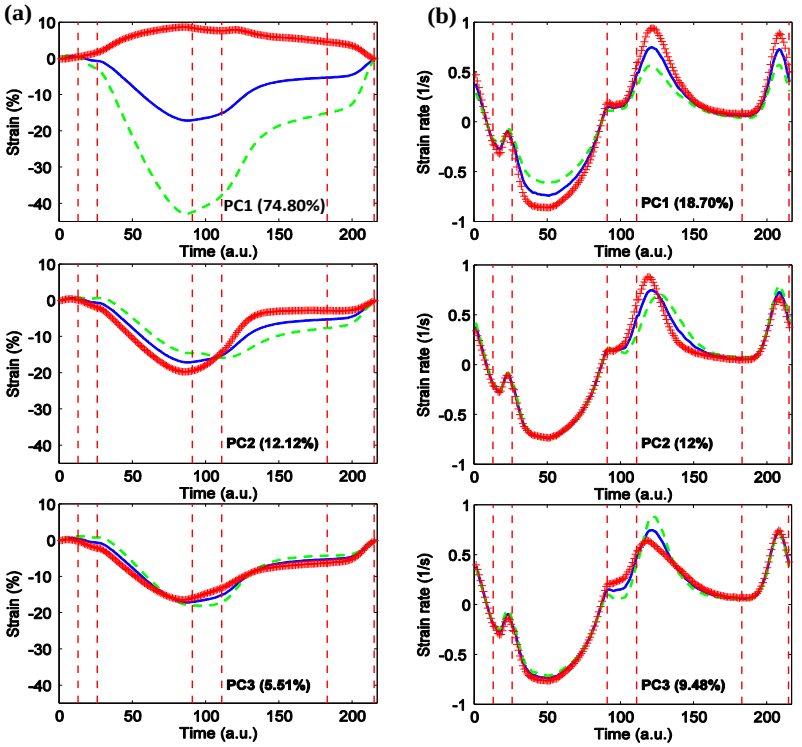


Figure 7.6: The first three modes of variation of the (a) strain and (b) strain rate PCs derived from the first PCA model. Mean curves plus (++red) and minus (- -green) 3σ for the strain PCs and 7σ for the strain rate PCs are presented. The percentage of data variation captured by each PC is also shown. Vertical dashed lines are the timing of the onset of the six mechanical phases of the cardiac cycle.

that some segments shorten less fast and their subsequent relaxation is less fast as well. The third PC captures early- and post-systolic variations among the curves. It shows that early-systolic lengthening is associated with post-systolic shortening which is known to be a hallmark of ischemia.

For the strain rate traces (Figure 7.6(b)), the first PC explains variations in overall strain rate amplitude and shows that high systolic strain rate is associated with high (early and late) diastolic strain rate and vice versa. The second PC models variations in timing and amplitude in the relaxation phase. It can be seen that a delayed and reduced curve in the early relaxation is associated with increased late diastolic strain rate. The third PC also explains variations in the early relaxation phase.

In Figure 7.6 the percentage of data variation explained by each PC is also shown. Two key observations can be made by comparing the percentages of the strain and strain rate's PCs:

1. the first three PCs of the strain curves account for 92.43% of the whole data variation while for the strain rate traces, 40.18% of the data variation is captured by the PCs,
2. the first PC in Figure 7.6(a) accounts for most of the strain curves variations while the PCs in Figure 7.6(b) have comparable contributions in describing variations of the strain rate curves.

These observations can be justified by the simpler temporal structure of the strain curves compared to those of the strain rate traces and imply that for modeling the strain curves, a smaller number of PCs is required.

The next PCs capture more complex temporal patterns of the deformation traces and making a proper patho-physiological interpretation of their variations is difficult. This is also the case for the PCs of the second PCA model (Figure 7.7) given the fact that this model is designed to describe temporal behavior of the deformation traces as well as spatial interactions between the LV segments simultaneously. Therefore, no interpretation of the second methodology's PCs is given

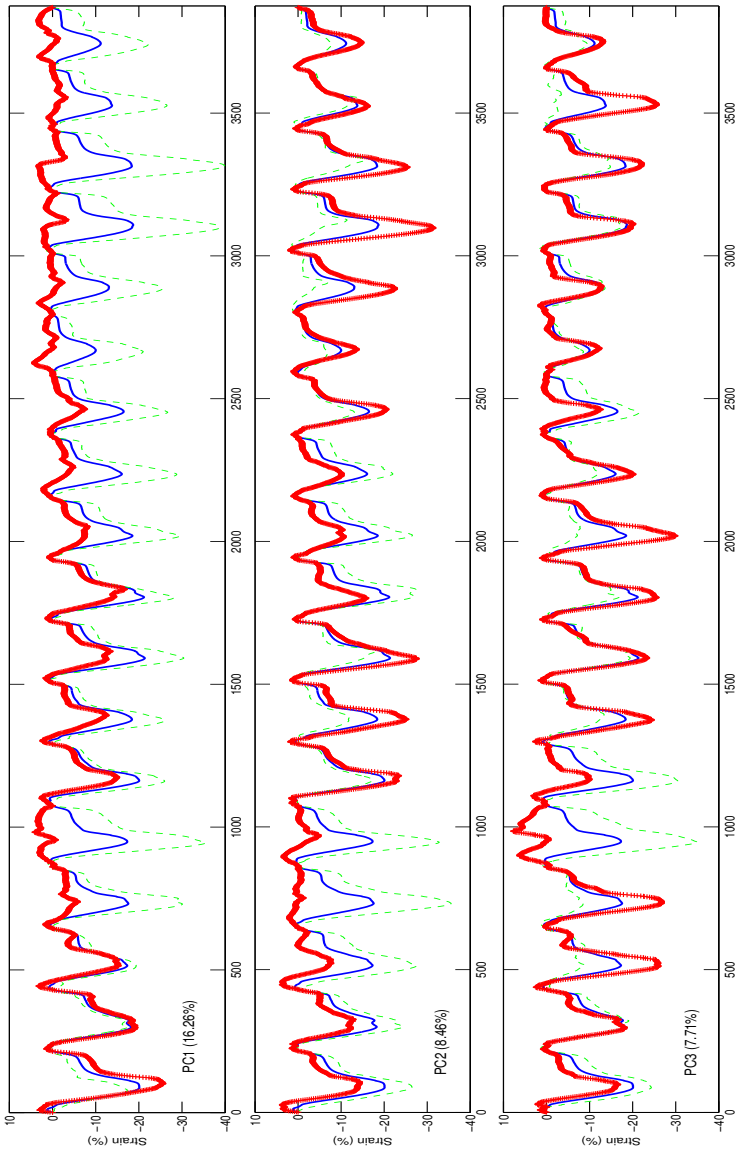


Figure 7.7: The first three modes of variation of the strain PCs derived from the second PCA model. Mean curves plus (+-red) and minus (- -green) 3σ of the PCs. The percentage of data variation captured by each PC is also shown.

Table 7.1: Number of the first and selected PCs from the two proposed methodologies that yielded to the best average classification performances.

	#First PCs	#Selected PCs
Methodology I		
Strain	10	3
Strain Rate	15	28
Methodology II		
Strain	21	13
Strain Rate	68	84

here but the usefulness of this model will be studied in the next section by considering the performances of the classification systems built based on its parameters.

7.4.2 Classification Results

For the PCA-derived features, different number of PCs were examined to have the optimal S-T representation of the LV function. From the first PCA model, the examined PCs per segment were taken from the set $\{5, 10, 15, 20, 25, 30\}$. The reason for limiting the maximum number of the tested PCs to 30 was the way of building the S-T representation in the first proposed methodology where the S-T feature vector's dimension was equal to the number of employed PCs multiplied by the number of the LV segments. Regarding the small number of subjects used in this study (i.e. 120), using large number of PCs increases the chance of overfitting. For the second PCA model, however, a wider range of the PCs could be examined because in the second methodology the computed PCs directly give the S-T representation of the LV function. Therefore, a subset of the PCs that captured percentages of data variation equal to $\{70\%, 80\%, 85\%, 90\%, 95\%, 99\%\}$ was examined.

Table 7.1 lists the number of first and selected PCs from each methodology that yielded to the best average classification perfor-

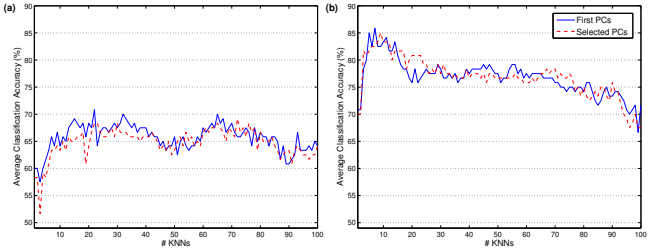


Figure 7.8: Average classification results on the (a) strain and (b) strain rate curves using the proposed S-T feature vectors in methodology I.

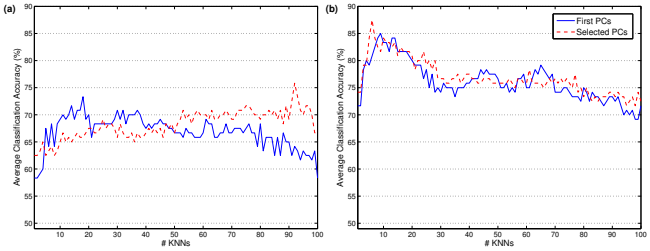


Figure 7.9: Average classification results on the (a) strain and (b) strain rate curves using the proposed S-T feature vectors in methodology II.

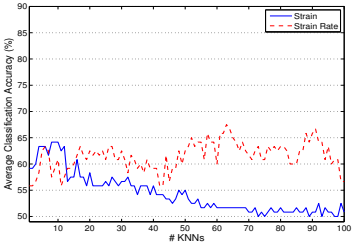


Figure 7.10: Average classification results on the strain and strain rate curves using the traditional features.

Table 7.2: Best average classification results (%) and their corresponding sensitivity and specificity values obtained by the S-T and traditional feature vectors.

	ACC	SE	SP
Methodology I			
First PCs			
Strain	70.83	78.33	63.33
Strain Rate	85.83	86.67	85
Selected PCs			
Strain	69.17	73.33	65
Strain Rate	85	85	85
Methodology II			
First PCs			
Strain	73.33	73.33	73.33
Strain Rate	85	86.67	83.33
Selected PCs			
Strain	75.83	73.33	78.33
Strain Rate	87.50	90	85
Traditional Features			
Strain	64.17	40	88.33
Strain Rate	67.50	80	55

mances. The average classification results corresponding to the PCs listed in Table 7.1 are demonstrated in Figure 7.8 and Figure 7.9 for the first and second methodology, respectively. The average classification outcomes obtained by the traditional features is shown in Figure 7.10. The best average classification outcomes and their corresponding sensitivity and specificity values obtained by the different feature vectors are listed in Table 7.2.

It can be seen from Figure 7.8, Figure 7.9 and Table 7.2 that the S-T features of methodology II could provide better results than those of

methodology I. The results also show that while by using the feature selection approach better classification performances were achieved than only using the first PCs for methodology II, methodology I could not take advantage of this technique. By considering the number of PCs that resulted in the best outcomes in Table 7.1, it can be realized that methodology II needs less number of PCs to give better results than methodology I. Note that, the number of PCs of methodology I, listed in Table 7.1, should be multiplied by the number of LV segments to account for the S-T feature vector's dimension.

Another key point that is highlighted by the results is that the strain rate traces convey significantly higher amount of information about the heart function than the strain curves where in both methodologies the S-T features extracted from the strain rate traces led to higher classification results than those of the strain curves. As shown in Table 7.1, the S-T feature vectors of the strain rate curves were generated by using significantly larger number of PCs compared to those of the strain curves. This observation confirms our expectation about the required number of PCs for modeling the behavior of the strain and strain rate traces regarding the complexity of their structure.

The obtained results show that both of the proposed S-T representations could outperform the traditional techno-markers for both strain and strain rate curves. This means that by exploring the whole cardiac cycle instead of considering one time point, considerably more information regarding the characteristics of the deformation traces can be achieved.

7.4.3 Cardiologist Decisions

Table 7.3 lists the cardiologist readings which were made based on the wall motion videos and all information mentioned in section 7.3.7. It shows that the cardiologist could make better decisions by using all available information sources than only using wall motion videos.

The cardiologist readings can be compared with the automatic classification results listed in Table 7.2 by recalling the difference between the number of subjects that were used in these two procedures. While

Table 7.3: Outcome of cardiologist readings (%) based on wall motion grayscale videos and having all three sources of information (i.e. wall motion videos, bull's eye plots and strain (rate) curves).

	ACC	SE	SP
Wall Motion	67.50	65.22	70.59
All Information	70	66.67	75

the cardiologist's readings are more accurate than the classification outcomes based on the traditional features, it could not provide better results than those of methodology II. Although the results of the visual assessment are comparable with the classification results achieved by the strain S-T features in methodology I, the strain rate S-T representation in methodology I yielded to significantly better performance than the cardiologist.

7.5 Conclusion

This study addressed the problem of modeling echocardiographic deformation curves for the automatic detection of MI. Two methodologies were proposed to model the S-T deformation patterns of the LV based on PCA. The PCA-derived S-T parameters were then used to build a classification system for the categorization of normal and infarcted hearts. Usefulness of the proposed S-T representations was confirmed through experimental evaluation where they could provide significantly better classification results than the traditional technomarkers (i.e. end-systolic strain and peak-systolic strain rate values) and visual assessments of an expert cardiologist. Classification results also demonstrated that the strain rate traces convey considerably more information about LV function than the strain curves and PCA is able to properly model this extra amount of discriminatory information content.

Chapter 8

Statistical Shape Modeling of the LV

This chapter is based on a paper presented in *Statistical Atlases and Computational Modeling of the Heart (STACOM) workshop, MICCAI 2015*, M. Tabassian, M. Alessandrini, P. Claes, L. De Marchi, D. Vandermeulen, G. Masetti and J. D'hooge, "Automatic detection of myocardial infarction through a global shape feature based on local statistical modeling"

8.1 Abstract

This study presents a local-to-global statistical approach for modeling the major components of left ventricular (LV) shape using its 3-D landmark representation. The rationale for dividing the LV into local areas is bi-fold: 1) to better identify abnormalities that lead to local shape remodeling and, 2) to decrease the number of shape variables by using a limited set of landmark points for an efficient statistical parametrization. Principal Component Analysis (PCA) is used for the statistical modeling of the local regions and subsets of the learned parameters that provide significant discriminatory information are taken from each local model in a feature selection stage. The selected local parameters are then concatenated to form a global representation of the LV and to train a classifier for differentiating between normal and infarcted LV shapes.

8.2 Introduction

Statistical shape analysis is a promising approach to model cardiac anatomy and to characterize myocardial abnormalities. The success of the point distribution model (PDM) [23] in describing anatomical structures of medical images makes it the basis of the majority of cardiac shape parametrization algorithms. These algorithms have been established using both linear methods (such as Principal Component Analysis (PCA) [39, 91, 104, 124] and Independent Component Analysis (ICA) [105]) and nonlinear techniques (such as kernel PCA [36]). One drawback of these techniques, however, is that they treat the shape globally. In addition to being computationally expensive due to the requirement of modeling a large number of variables, a global approach may fail to characterize abnormalities that affect small regions of the myocardium. An alternative approach is to learn local statistical shape components and then merge their results to describe the global shape as a poly-local model. A recent and well-established example of such framework is presented in [123]. It is based on utilizing local shape descriptors, but not landmark points as suggested in PDM,

and employing a manifold learning technique called ISOMAP [111] for dimension reduction.

Inspired by [123], a local statistical shape modeling approach based on PDM and PCA [61] is presented in this study to characterize major components of LV shapes. The rationale of utilizing PCA in our framework is as follows: 1) PCA implementation involves simple steps and its parameters can be efficiently computed and, 2) it allows to visualize major modes of data variation. The latter property could be of particular interest to study the relation between the parameters of the statistical model and the patho-physiology of the heart.

The main contribution of this study compared to the framework presented in [123] is the way that the local statistical information are incorporated in the classification phase. In [123], an independent classifier was built with the parameters of each local model and the classifiers decisions were fused using majority voting. Independent treatment of the local models parameters could degrade the capability of the combined classification model in dealing with abnormalities that affect small regions of the heart. Here, we propose to create an alternative local-to-global representation of the LV shape components by concatenating the parameters of the local models and then building a classifier with the obtained feature vector. Having the advantage of encoding global shape parameters of the LV, the spatial relation between the local zones is taken into consideration using this technique. Explicit usage of the local statistical parameters can also create distinct areas in the global feature space. This property enables a classification system to better characterize abnormalities that mostly affect small regions of the myocardium.

8.3 Materials and Method

Figure 8.1 represents global and local architectures that were implemented in this study for the statistical modeling and classification of LV shapes. In the local architecture (Figure 8.1(b)), the LV was divided into non-overlapping regions of interest (ROI) and an independent PCA model was built with the local shapes belonging to each

ROI. By taking a subset of the learned statistical parameters, two different classification schemes were examined. In addition, a global PCA model (Figure 8.1(a)) was also built to benchmark the performances of the local PCA models.

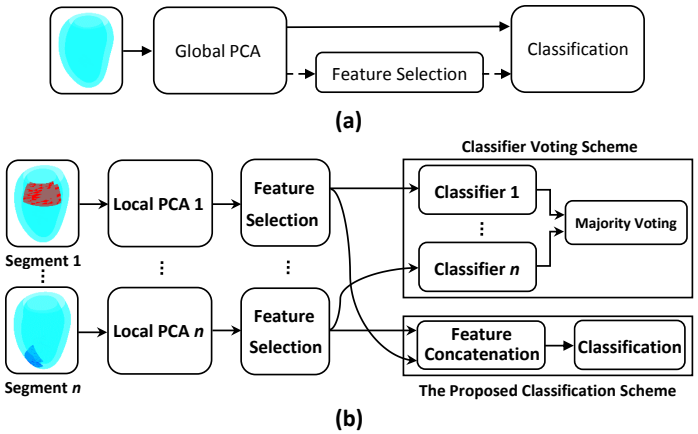


Figure 8.1: Architectures of the implemented statistical frameworks. (a) A PCA model is constructed with whole LV shapes. Then, a subset of the first PCs (solid line) or selected PCs (dashed line) is used for training a classifier. (b) Independent PCA models are built with the segments of LV shapes and subsets of the selected PCs are used in the classification phase. In the classifier voting scheme, independent classifiers are trained with the selected PCs of the local models and the final decision is made by the majority voting rule. In the proposed scheme, one classifier is trained with a feature vector that is obtained by concatenating the local models parameters.

8.3.1 Data and Preprocessing

A data set of 100 healthy volunteers and 100 patients with myocardial infarction (MI) from the MESA [8] and DETERMINE studies [62] respectively, was used in our experiments. These data sets are part of the Cardiac Atlas Project (CAP, www.cardiacatlas.org) [38] and contain cardiovascular magnetic resonance (CMR) images. Endocardial

and epicardial shapes at end-diastole (ED) and end-systole (ES) are represented with their corresponding Cartesian sets of landmark points in magnet coordinates. It has been demonstrated in [124] that a PCA model built with shape vectors at ED and ES could provide better outcomes than its counterparts that were constructed with only ED or ES shapes. Therefore, the shapes of both ED and ES cardiac phases were used for the implementation of the global and local PCA models in the current study.

All shapes were aligned by making use of the generalized Procrustes superimposition method [95]. As suggested in [124], for building the global PCA model the alignment phase has been performed by eliminating position and orientation differences but preserving scale variations as ventricular size has a predictive value for diagnosing MI. For constructing the local PCA models, however, scale variations were also removed in the alignment procedure.

8.3.2 Statistical Modeling

Both global and local PCA models were learned using the data of the healthy volunteers to capture major modes of normal shape variations. For building the local PCA models, small, medium and large ROI sizes were examined which respectively encompassed 4, 8 and 16 faces of the 3-D meshes in both the circumferential and longitudinal directions where the full LV mesh was composed of 32×32 faces. Note that the landmark points in each ROI were consistent across the subjects.

8.3.3 Feature Selection

The ultimate goal of the presented framework is the accurate categorization of the normal and infarcted LV shapes. This requires that the statistical parameters taken for training a classification system provide significant discriminatory information. Traditionally, data are projected onto the subspace spanned by the first principal components (PCs) to retain most of the variation in the original variables. However, it is possible that some of the PCs with low contribution in the

data variation contain relevant discriminatory information. As such, the P-metric method [55], presented in section 1.4, was used in our framework to select relevant PCs. The computed PCs were sorted based on their P-metric values in descending order and a subset of the first selected PCs were used in the classification stage.

8.3.4 Classification

Figure 8.1(b) illustrates two classification schemes that were trained with the parameters of the local PCA models. The first scheme uses the strategy proposed in [123] while the second one works based on the idea of concatenating the local PCA models parameters. Both methodologies were implemented by making use of a subset of PCs that had been chosen in the feature selection stage. Note that, since the local PCA models were trained independently, the selected PCs for each model can be different from the others. SVM [25, 117] with a linear kernel was used as classifier in both global and local models. Classification outcomes were obtained using 10-fold cross-validation. Hereto, data vectors were randomly divided into 10 equal-size folds such that each fold had the same number of patterns from each class. Classifiers were trained with the first nine folds and tested with the last one. This procedure was repeated 10 times so that all folds were used for training and testing the classifiers.

8.4 Results

Average classification resuls obtained with the global PCA model are shown in Figure 8.2. While the best result of the classifier trained with the first PCs (Figure 8.2(a)) was achieved by preserving 99% of the data variation (the first 60 PCs), training the classifier with the selected PCs (Figure 2(b)) yielded better performance with considerably less number of features (6 PCs).

Figure 8.3 illustrates the average classification accuracies achieved by the local PCA models. It can be seen that all local models could

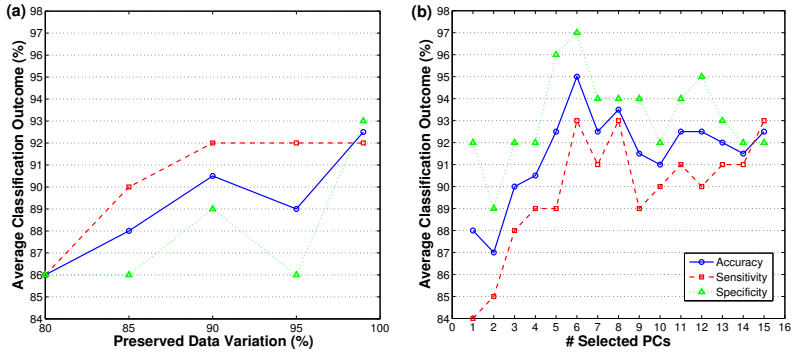


Figure 8.2: Average classification outcomes (%) with the global PCA model. (a) Training the classifier with a subset of the first PCs or (b) with a subset of the selected PCs.

provide significantly higher classification results than the global ones. The best classification results of the global and local PCA models along with their corresponding sensitivity and specificity values are listed in Tables 8.1 and 8.2, respectively.

In order to give insight into the characteristics of the selected PCs that enable a classifier to discriminate between the normal and infarcted LV shapes, the first five selected modes of variation of the global PCA model, which were observed constantly across the 10 folds, are visualized in Figure 8.4. Note that the reason for visualizing the global PCs is that they are easier to interpret than the local PCs.

8.5 Discussion

8.5.1 Global versus Local Statistical Modeling

It has been demonstrated that by using a limited number of the landmark points, a local framework is able to provide better statistical description of LV shapes than its global counterpart. The performance

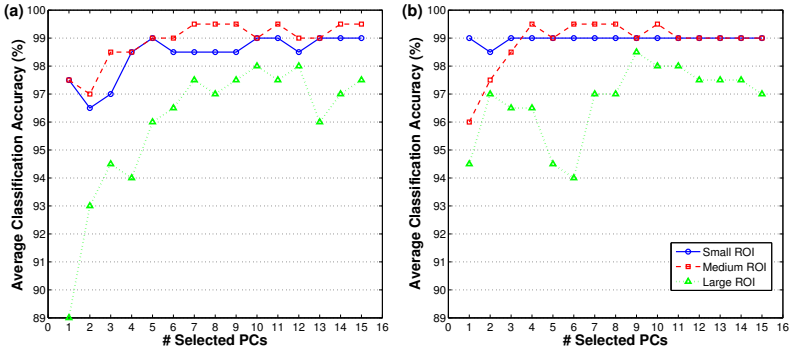


Figure 8.3: Average classification accuracies (%) obtained with the selected PCs of the local PCA models and utilizing (a) the classifier voting scheme and (b) the proposed classification framework.

of the local structures, however, depends on the ROI size and tuning this parameter needs a proper compromise between the statistical significance and number of landmark points. Although the performance of the local classification schemes in detecting MI is comparable, direct usage of the local models parameters and considering the spatial relation between the LV segments would enable the proposed scheme to properly deal with different abnormalities that affect small regions of the myocardium.

As shown in Figure 8.3, the favorable results of the classifier voting and the proposed scheme were obtained by using a few number of selected PCs per local PCA model which can be explained by the following reasons: 1) local regions have less modes of variation than the whole LV and their statistical modeling needs less number of components as well, 2) as shown in Figure 8.2 (b), selected PCs could provide considerable discriminatory information for the local classification systems.

Table 8.1: Best average classification accuracy results (%) and their corresponding sensitivity and specificity values obtained with the global PCA model (typed in bold). The (min,max) ranges of the obtained outcomes are also presented.

	ACC	SE	SP
First PCs	**92.5** (80,100)	**92** (80,100)	**93** (80,100)
Selected PCs	**95** (80,100)	**93** (80,100)	**97** (80,100)

Table 8.2: Best average classification accuracy results (%) and their corresponding sensitivity and specificity values obtained with the local PCA models (typed in bold). The (min,max) ranges of the obtained outcomes are also presented.

	ACC	SE	SP
Classification Voting Scheme			
Small ROI	**99** (95,100)	**98** (90,100)	**100** (100,100)
Medium ROI	**99.5** (95,100)	**99** (90,100)	**100** (100,100)
Large ROI	**98** (90,100)	**98** (90,100)	**98** (90,100)
The proposed Scheme			
Small ROI	**99** (95,100)	**98** (90,100)	**100** (100,100)
Medium ROI	**99.5** (95,100)	**99** (90,100)	**100** (100,100)
Large ROI	**98.5** (95,100)	**97** (90,100)	**100** (100,100)

8.5.2 Feature Selection Utility

The obtained results confirm the suitability of the feature selection strategy where training a classifier with a small group of the selected PCs could significantly enhance the performance of the same classifier that was trained with a much larger subset of the first PCs. To shed further light on the utility of the feature selection method and

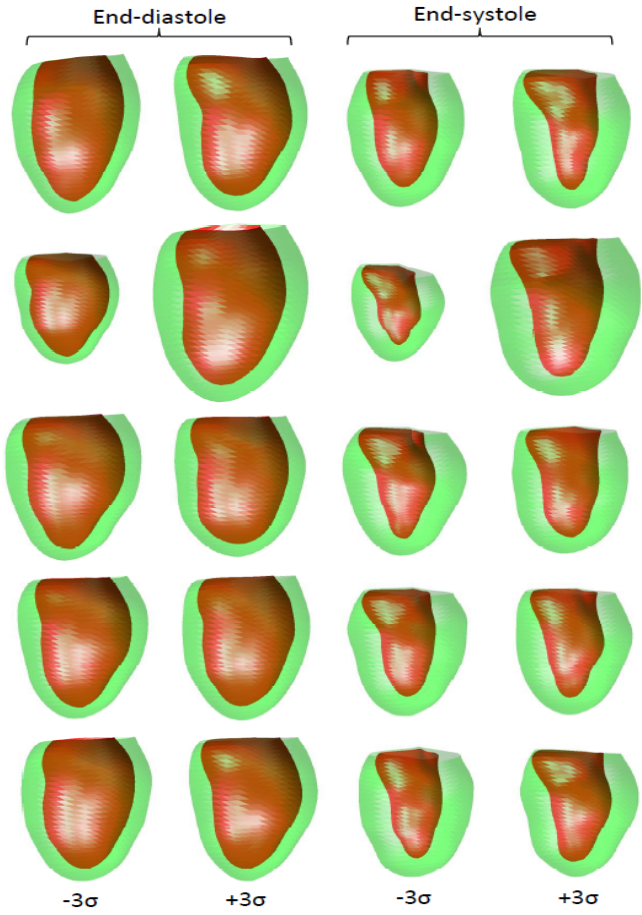


Figure 8.4: Variations of the first five selected PCs of the global PCA model at end-diastole and end-systole. From top, PCs 18, 1, 20, 42 and 15 are the selected modes of variation.

statistical shape modeling with PCA, some patho-physiological interpretations of the selected PCs belong to the global PCA model (shown in Figure 8.4) are given in the following.

The left- and righ-hand sides of each LV mode in Figure 8.4 correspond to the anteroseptal and inferolateral walls, respectively. The first selected mode (PC 18) describes variations in the curvature of the anterior wall. This PC might have been selected also due to possible difference in contouring convention of the left ventricular outflow tract (LVOT) in the MESA and DETERMINE trials. The second selected mode (PC 1) explains variations in the LV size. Blunting of the apex and variation of the inferior wall curvature is described by the third selected mode (PC 20). The forth selected mode (PC 42) is associated with the end-systolic variations in the curvatures of the inferior region and the anterior wall. Finally, the fifth selected mode (PC 15) captures the rightward shifting of the apex and variations in the inferior region.

The above-mentioned patho-physiological interpretations are mostly based on the evidences presented in [31, 83] and are associated with the process of the LV remodeling due to anterior MI. Although the DETERMINE study involves patients with different types of the MI, it is well-known that coronary artery disease occurs most commonly in the left anterior descending (LAD) coronary artery. Therefore, interpretation of the selected PCs based on the findings of [31, 83] might be valid for the majority of the subjects in this study.

8.6 Conclusion

A statistical framework has been established based on local PCA models to characterize major modes of LV shape variation. Although local statistical modeling could bring favorable advantages over global parametrization, the adopted strategy for associating the local models parameters plays a key role in obtaining an efficient local-to-global shape characterization. We hypothesized that the concatenation of the local models parameters would lead to such efficient characterization. Parameters of each local model were selected based on their significance in discriminating normal and infarcted shapes. Classification results confirmed the superiority of the proposed statistical framework over the global model. They also approved the suitability of the feature selection strategy where utilizing a few number of selected PCs could

yield high classification results.

Chapter 9

Conclusions and Future Perspectives

9.1 Summary and General Conclusions

This thesis was concerned with developing new computerized ultrasound-based systems for the diagnosis of PCa and MI. The advantages of ultrasound imaging, like safety, portability, being real-time and cheap, allow the proposed CAD systems to be widely used in the clinical practice. In the following, the major contributions made in this thesis are reviewed.

In Chapter 3, a label assignment approach was introduced to deal with prostate biopsy samples with uncertain labels. The accepted label uncertainties were then managed in the framework of the D-S theory of evidence. By using this approach, the size of the training data can be increased, which is of great advantage for building a solid classification system, and data wasting is also avoided.

The ICA method was used in Chapter 4 to learn features from the RF backscattered signals to diagnose PCa. By using ICA for learning features from the RF signals, less restrictive assumptions about the interaction between ultrasound waves and tissue are made compared to the traditional feature extraction. This property can bring high generalization power to a CAD system. In this chapter, an ensemble system for learning ICA dictionaries and classifying four different classes of backscatterers was also proposed. It has been shown that by using an ensemble-based methodology for dictionary learning, (near-) optimal parameters of the ICA model can be found and favorable classification results could be achieved by merging the classification outputs constructed with the employed ICA dictionaries.

In Chapter 6, PCA was used to model major temporal patterns of segmental strain (rate) traces acquired from the LV. A classification system was then trained to categorize normal and infarcted segments. The proposed method could yield better results than a classifier trained with some common traditional techno-markers like end-systolic strain and peak-systolic strain rate values. Moreover, the diagnostic procedure was automatic and could provide objective evaluation of the segmental LV function.

PCA was also used in Chapter 7 with the aim of modeling the S-T

characteristics of the LV deformation function. Two different PCA-based methods were proposed to differentiate normal volunteers from the patients with acute MI. The proposed models were contrasted with S-T representations built from the traditional techno-markers and also with the readings of an expert cardiologist. The obtained results confirmed the superiority of the proposed PCA-based methods over the other employed approaches.

Finally, a local-to-global statistical approach based on PCA was introduced in Chapter 8 to model shape parameters of the LV to detect MI. Statistical modeling was performed locally by building independent PCA models on non-overlapping ROIs from the LV shapes. The CAD system took advantage of a feature selection approach to sort PCs of each PCA model based on their discrimination information. Small subsets of the sorted PCs were then taken from the local models and by connecting them together, a global representation of the LV shapes was achieved. The proposed scheme could yield to higher classification results than a global shape model learned also by PCA. The usefulness of the utilized feature selection approach and also PCA, as the statistical modeling approach, were demonstrated by visualizing some of the first selected PCs of the global PCA model and illuminating their links with patho-physiology.

9.2 Future Perspectives

9.2.1 Using Nonlinear statistical techniques

In this thesis, PCA and ICA have been used for data modeling. Both of these methods, however, explore linear relationships between data variables and do not take into account their nonlinear dependencies. As an alternative to these methods, performances of some nonlinear statistical methods like kernel PCA [100], nonlinear ICA [53] or ISOMAP manifold learning [111] can be investigated.

9.2.2 Applying the Proposed ICA Dictionary Learning Method to IQ or B-mode images

The proposed ensemble dictionary learning approach presented in Chapter 4 has been developed to extract feature from RF signals. In some applications, however, RF data is not easily available or the diagnostic procedure is mainly based on the other types of ultrasound images. The generality of the proposed dictionary learning method allows it to also be applied to the other widely-used ultrasound images like B-mode.

9.2.3 Combining Tissue, Function and Shape Attributes

Some of the cardiovascular diseases (e.g. MI) change tissue [121], function and shape characteristics of the heart. Therefore, an optimal approach to characterize myocardial abnormalities is to combine the tissue, function and shape attributes of the heart. In Chapters 4, 7 and 8 of this thesis, three schemes were presented for modeling tissue structure, LV function and LV shape patterns, respectively. Having the parameters derived by these schemes at hand, the next step could be to merge them for a better assessment of the heart abnormalities.

9.2.4 Using Other Feature Selection Methods

In Chapters 7 and 8, the P-metric method was used for selecting discriminating features. This approach, which falls into the category of filter feature selection methods, considers means and standard deviations of the classes in the subspace of each feature to determine its relevance. Other members of the filter feature selection family, which use different types of criteria for sorting the relevant features, can be employed as well. Furthermore, the wrapper feature selection techniques [46], which evaluate the usefulness of the features using a learning machine, can be used in the structure of the proposed CAD systems.

9.2.5 Using Characteristics of the Patients

Physical and clinical characteristics of the patients can also be incorporated into the proposed CAD systems to enhance their performances. In addition to general characteristics like sex, age, height and weight, some other information that has stronger correlation with pathology can be used as well. This can be the results of PSA and DRE tests for diagnosing PCa. For the CAD systems designed for detecting abnormalities of the heart, utilizing some clinical indexes like systolic and diastolic blood pressures, ejection fraction and presence or absence of diabetes can be advantageous to the diagnostic scheme. Some of the behavioral characteristics related to the cardiac diseases like smoking and alcohol usage can be taken into account as well.

9.2.6 Applying the Proposed Methods to Other Pathologies

The proposed CAD systems can be applied to other pathologies as well. As an example, the methods proposed to detect MI can also be used to characterize hypertension as this abnormality affects both motion and shape patterns of the LV. Given the fact that the proposed ICA-based methods have been designed to address feature extraction in medical ultrasound, they can be applied to a wide range of pathologies as long as enough data for training the ICA model are available. The approach presented in Chapter 3 can be used to deal with those group of CAD problems that contain data with uncertain or inaccurate labels. This issue can be seen in clinical practice where the labels associated to the data are not reliable/accurate due to the measurement noise or the difficulty of assigning crisp class labels to the acquired measurements.

Bibliography

[1] M. Aharon, M. Elad, and A. Bruckstein. "SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation". In: *IEEE Trans. Signal Process.* 54.11 (Nov. 2006), pp. 4311–4322. DOI: 10.1109/TSP.2006.881199.

[2] F. Aoued, E. Eroglu, L. Herbots, F. Rademakers, and J. D'hooge. "A statistical model-based approach for the detection of abnormal cardiac deformation". In: *IEEE Ultrason. Symp. 2005.* Vol. 1. IEEE, 2005, pp. 512–515. DOI: 10.1109/ULTSYM.2005.1602903.

[3] M. Bartlett, J. Movellan, and T. Sejnowski. "Face recognition by independent component analysis". In: *IEEE Trans. Neural Networks* 13.6 (Nov. 2002), pp. 1450–1464. DOI: 10.1109/TNN.2002.804287.

[4] G. Baudat and F. Anouar. "Generalized Discriminant Analysis Using a Kernel Approach". In: *Neural Comput.* 12.10 (Oct. 2000), pp. 2385–2404. DOI: 10.1162/089976600300014980.

[5] Y. Bengio, A. Courville, and P. Vincent. "Representation Learning: A Review and New Perspectives". In: *IEEE Trans. Pattern Anal. Mach. Intell.* 35.8 (Aug. 2013), pp. 1798–1828. DOI: 10.1109/TPAMI.2013.50.

[6] O. Bernard, J. D'Hooge, and D. Friboulet. "Statistics of the radio-frequency signal based on K distribution with application to echocardiography". In: *IEEE Trans. Ultrason. Ferroelectr.*

Freq. Control 53.9 (2006), pp. 1689–1694. DOI: 10.1109/TUFFC.2006.1678198.

[7] O. Bernard, B. Touil, J. Dfffdfffdhooge, and D. Friboulet. "Statistical Modeling of the Radio-Frequency Signal for Partially- and Fully-Developed Speckle Based on a Generalized Gaussian Model with Application to Echocardiography". In: *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 54.10 (Oct. 2007), pp. 2189–2194. DOI: 10.1109/TUFFC.2007.515.

[8] D. E. Bild. "Multi-Ethnic Study of Atherosclerosis: Objectives and Design". In: *Am. J. Epidemiol.* 156.9 (Nov. 2002), pp. 871–881. DOI: 10.1093/aje/kwf113.

[9] C. M. Bishop. *Pattern recognition and machine learning*. springer, 2006.

[10] A. Blum and T. Mitchell. "Combining labeled and unlabeled data with co-training". In: *Proc. Elev. Annu. Conf. Comput. Learn. theory - COLT' 98*. New York, New York, USA: ACM Press, 1998, pp. 92–100. DOI: 10.1145/279943.279962.

[11] L. Breiman. "Bagging predictors". In: *Mach. Learn.* 24.2 (Aug. 1996), pp. 123–140. DOI: 10.1007/BF00058655.

[12] C. J. Burges. "A Tutorial on Support Vector Machines for Pattern Recognition". In: *Data Min. Knowl. Discov.* 2.2 (1998), pp. 121–167. DOI: 10.1023/A:1009715923555.

[13] V. D. Calhoun, J. Liu, and T. Adali. "A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data". In: *Neuroimage* 45.1 (Mar. 2009), S163–S172. DOI: 10.1016/j.neuroimage.2008.10.057.

[14] R. a. Castellino. "Computer aided detection (CAD): an overview". In: *Cancer Imaging* 5.1 (2005), pp. 17–19. DOI: 10.1102/1470-7330.2005.0018.

[15] W. J. Catalona et al. "Comparison of digital rectal examination and serum prostate specific antigen in the early detection of prostate cancer: results of a multicenter clinical trial of 6,630 men." In: *J. Urol.* 151.5 (1994), pp. 1283–90.

[16] M. D. Cerqueira et al. “Standardized Myocardial Segmentation and Nomenclature for Tomographic Imaging of the Heart: A Statement for Healthcare Professionals From the Cardiac Imaging Committee of the Council on Clinical Cardiology of the American Heart Association”. In: *Circulation* 105.4 (Jan. 2002), pp. 539–542. DOI: 10.1161/hc0402.102975.

[17] J. Cheng, Q. Liu, H. Lu, and Y.-W. Chen. “Ensemble learning for independent component analysis”. In: *Pattern Recognit.* 39.1 (Jan. 2006), pp. 81–88. DOI: 10.1016/j.patcog.2005.06.018.

[18] G. Cincotti, G. Loi, and M. Pappalardo. “Frequency decomposition and compounding of ultrasound medical images with wavelet packets”. In: *IEEE Trans. Med. Imaging* 20.8 (2001), pp. 764–771. DOI: 10.1109/42.938244.

[19] P. Clarysse, M. Han, P. Croisille, and I. Magnin. “Exploratory analysis of the spatio-temporal deformation of the myocardium during systole from tagged MRI”. In: *IEEE Trans. Biomed. Eng.* 49.11 (Nov. 2002), pp. 1328–1339. DOI: 10.1109/TBME.2002.804587.

[20] P. Claus, J. D’hooge, T. Langeland, B. Bijnens, and G. Sutherland. “SPEQLE (Software package for echocardiographic quantification LEuven) an integrated approach to ultrasound-based cardiac deformation quantification”. In: *Comput. Cardiol.* IEEE, 2002, pp. 69–72. DOI: 10.1109/CIC.2002.1166709.

[21] R. Clements, O. Aideyan, G. Griffiths, and W. Peeling. “Side effects and patient acceptability of transrectal biopsy of the prostate”. In: *Clin. Radiol.* 47.2 (Feb. 1993), pp. 125–126. DOI: 10.1016/S0009-9260(05)81188-2.

[22] P. Comon. “Independent component analysis, A new concept?” In: *Signal Processing* 36.3 (Apr. 1994), pp. 287–314. DOI: 10.1016/0165-1684(94)90029-9.

[23] T. Cootes, C. Taylor, D. Cooper, and J. Graham. *Active Shape Models-Their Training and Application*. Jan. 1995. DOI: 10.1006/cviu.1995.1004.

[24] T. Cover and P. Hart. “Nearest neighbor pattern classification”. In: *IEEE Trans. Inf. Theory* 13.1 (Jan. 1967), pp. 21–27. DOI: 10.1109/TIT.1967.1053964.

[25] N. Cristianini and J. Shawe-Taylor. *An introduction to Support Vector Machines*. Vol. 47. 2. 2000, pp. 1–15.

[26] S. Curran, O. Akin, A. M. Agildere, J. Zhang, H. Hricak, and J. Rademaker. “Endorectal MRI of Prostatic and Periprostatic Cystic Lesions and Their Mimics”. In: *Am. J. Roentgenol.* 188.5 (May 2007), pp. 1373–1379. DOI: 10.2214/AJR.06.0759.

[27] A. P. Dempster. “Upper and lower probabilities induced by a multivalued mapping”. In: *Ann. Math. Stat.* 38.2 (1967), pp. 325–339. DOI: 10.2307/2239146.

[28] W. Desmet et al. “High-dose intracoronary adenosine for myocardial salvage in patients with acute ST-segment elevation myocardial infarction”. In: *European heart journal* 32.7 (2011), pp. 867–877. DOI: 10.1093/eurheartj/ehq492.

[29] J. D’hooge, B. Bijnens, J. Thoen, F. Van de Werf, G. Sutherland, and P. Suetens. “Echocardiographic strain and strain-rate imaging: a new tool to study regional myocardial function”. In: *IEEE Trans. Med. Imaging* 21.9 (Sept. 2002), pp. 1022–1030. DOI: 10.1109/TMI.2002.804440.

[30] J. D’hooge et al. “Regional Strain and Strain Rate Measurements by Cardiac Ultrasound: Principles, Implementation and Limitations”. In: *Eur. J. Echocardiogr.* 1.3 (Sept. 2000), pp. 154–170. DOI: 10.1053/euje.2000.0031.

[31] M. Di Donato et al. “Left ventricular geometry in normal and post-anterior myocardial infarction patients: sphericity index and new conicity index comparisons”. In: *Eur. J. Cardio-Thoracic Surg.* 29 (Apr. 2006), pp. 225–230. DOI: 10.1016/j.ejcts.2006.03.002.

[32] Di Lai, N. Rao, C.-h. Kuo, S. Bhatt, and V. Dogra. “Independent Component Analysis Applied to Ultrasound Speckle Texture Analysis and Tissue Characterization”. In: *2007 29th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* Vol. 2007. IEEE,

Aug. 2007, pp. 6523–6526. DOI: 10.1109/IEMBS.2007.4353854.

[33] K. Doi. "Computer-aided diagnosis in medical imaging: Historical review, current status and future potential". In: *Comput. Med. Imaging Graph.* 31.4-5 (June 2007), pp. 198–211. DOI: 10.1016/j.compmedimag.2007.02.002.

[34] S. A. Dudani. "The Distance-Weighted k-Nearest-Neighbor Rule". In: *IEEE Trans. Syst. Man. Cybern.* SMC-6.4 (Apr. 1976), pp. 325–327. DOI: 10.1109/TSMC.1976.5408784.

[35] V. Dutt and J. F. Greenleaf. "Ultrasound Echo Envelope Analysis Using a Homodyned K Distribution Signal Model". In: *Ultrason. Imaging* 16.4 (Oct. 1994), pp. 265–287. DOI: 10.1177/016173469401600404.

[36] S. Faghih Roohi and R. Aghaeizadeh Zoroofi. "4D statistical shape modeling of the left ventricle in cardiac MR images". In: *Int. J. Comput. Assist. Radiol. Surg.* 8.3 (May 2013), pp. 335–351. DOI: 10.1007/s11548-012-0787-1.

[37] E. J. Feleppa et al. "Recent Developments in Tissue-Type Imaging (TTI) for Planning and Monitoring Treatment of Prostate Cancer". In: *Ultrason. Imaging* 26.3 (July 2004), pp. 163–172. DOI: 10.1177/016173460402600303.

[38] C. G. Fonseca et al. "The Cardiac Atlas Project–an imaging database for computational modeling and statistical atlases of the heart". In: *Bioinformatics* 27.16 (Aug. 2011), pp. 2288–2295. DOI: 10.1093/bioinformatics/btr360.

[39] A. F. Frangi, D. Rueckert, J. A. Schnabel, and W. J. Niessen. "Automatic construction of multiple-object three-dimensional statistical shape models: Application to cardiac modeling". In: *IEEE Trans. Med. Imaging* 21.9 (2002), pp. 1151–1166. DOI: 10.1109/TMI.2002.804426.

[40] F. Galluzzo, N. Testoni, L. De Marchi, N. Speciale, and G. Masetti. "Improving Prostate Biopsy Protocol with a Computer Aided Detection Tool Based on Semi-supervised Learning". In:

Prostate Cancer Imaging. 2011, pp. 109–120. DOI: 10.1007/978-3-642-23944-1{_}12.

[41] M. J. Gangeh, A. Sadeghi-Naini, M. S. Kamel, and G. J. Czarnota. "Assessment of cancer therapy effects using texton-based characterization of quantitative ultrasound parametric images". In: *2013 IEEE 10th Int. Symp. Biomed. Imaging*. IEEE, Apr. 2013, pp. 1372–1375. DOI: 10.1109/ISBI.2013.6556788.

[42] Z. Ge and Z. Song. "Performance-driven ensemble learning ICA model for improved non-Gaussian process monitoring". In: *Chemom. Intell. Lab. Syst.* 123 (2013), pp. 1–8. DOI: 10.1016/j.chemolab.2013.02.001.

[43] G. Georgiou and F. Cohen. "Tissue characterization using the continuous wavelet transform. I. Decomposition method". In: *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 48.2 (Mar. 2001), pp. 355–363. DOI: 10.1109/58.911718.

[44] G. Georgiou, F. Cohen, C. Piccoli, F. Forsberg, and B. Goldberg. "Tissue characterization using the continuous wavelet transform. II. Application on breast RF data". In: *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 48.2 (Mar. 2001), pp. 364–373. DOI: 10.1109/58.911719.

[45] D. F. Gleason. "Histologic grading of prostate cancer: A perspective". In: *Hum. Pathol.* 23.3 (Mar. 1992), pp. 273–279. DOI: 10.1016/0046-8177(92)90108-F.

[46] I. Guyon and A. Elisseeff. "An introduction to variable and feature selection". In: *J. Mach. Learn. Res.* 3 (2003), pp. 1157–1182. DOI: 10.1162/153244303322753616.

[47] A. Heimdal, A. Støylen, H. Torp, and T. Skjærpe. "Real-Time Strain Rate Imaging of the Left Ventricle by Ultrasound". In: *J. Am. Soc. Echocardiogr.* 11.11 (Nov. 1998), pp. 1013–1019. DOI: 10.1016/S0894-7317(98)70151-8.

[48] L. Herbots et al. "Improved regional function after autologous bone marrow-derived stem cell transfer in patients with acute myocardial infarction: a randomized, double-blind strain rate

imaging study". In: *Eur. Heart J.* 30.6 (Aug. 2008), pp. 662–670. DOI: 10.1093/eurheartj/ehn532.

[49] G. E. Hinton. "Reducing the Dimensionality of Data with Neural Networks". In: *Science* 313.5786 (July 2006), pp. 504–507. DOI: 10.1126/science.1127647.

[50] K. K. Hodge, J. E. McNeal, M. K. Terris, and T. a. Stamey. "Random systematic versus directed ultrasound guided transrectal core biopsies of the prostate." In: *J. Urol.* 142.1 (July 1989), pp. 71–4.

[51] A. Hyvärinen. "Fast and robust fixed-point algorithms for independent component analysis". In: *IEEE Trans. Neural Networks* 10.3 (May 1999), pp. 626–634. DOI: 10.1109/72.761722.

[52] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent component analysis*. John Wiley & Sons, 2001.

[53] A. Hyvärinen and P. Pajunen. "Nonlinear independent component analysis: Existence and uniqueness results". In: *Neural Networks* 12.3 (Apr. 1999), pp. 429–439. DOI: 10.1016/S0893-6080(98)00140-3.

[54] M. F. Insana, T. J. Hall, and J. L. Fishback. "Identifying Acoustic Scattering Sources in Normal Renal Parenchyma from the Anisotropy in Acoustic Properties". In: *Ultrasound Med. Biol.* 17.6 (1991), pp. 613–626.

[55] I. Inza, P. Larrañaga, R. Blanco, and A. J. Cerrolaza. "Filter versus wrapper gene selection approaches in DNA microarray domains". In: *Artif. Intell. Med.* 31.2 (June 2004), pp. 91–103. DOI: 10.1016/j.artmed.2004.01.007.

[56] A. Jain, P. Duin, and Jianchang Mao. "Statistical pattern recognition: a review". In: *IEEE Trans. Pattern Anal. Mach. Intell.* 22.1 (2000), pp. 4–37. DOI: 10.1109/34.824819.

[57] F. Jamal et al. "Can changes in systolic longitudinal deformation quantify regional myocardial function after an acute infarction? An ultrasonic strain rate and strain study". In: *J. Am. Soc. Echocardiogr.* 15.7 (July 2002), pp. 723–730. DOI: 10.1067/mje.2002.118913.

[58] S. Janssens et al. “Autologous bone marrow-derived stem-cell transfer in patients with ST-segment elevation myocardial infarction: double-blind, randomised controlled trial”. In: *The Lancet* 367.9505 (2006), pp. 113–121. DOI: 10.1016/S0140-6736(05)67861-0.

[59] R. Jasaityte and J. D’hooge. “Strain rate imaging: fundamental principles and progress so far”. In: *Imaging Med.* 2.5 (Oct. 2010), pp. 547–563. DOI: 10.2217/iim.10.41.

[60] Jiao Wang, Si-wei Luo, and Xian-hua Zeng. “A random subspace method for co-training”. In: *2008 IEEE Int. Jt. Conf. Neural Networks (IEEE World Congr. Comput. Intell.* IEEE, June 2008, pp. 195–200. DOI: 10.1109/IJCNN.2008.4633789.

[61] I. Jolliffe. “Principal Component Analysis”. In: *Wiley StatsRef Stat. Ref. Online*. Chichester, UK: John Wiley & Sons, Ltd, Sept. 2014. DOI: 10.1002/9781118445112.stat06472.

[62] A. H. KADISH et al. “Rationale and Design for the Defibrillators to Reduce Risk by Magnetic Resonance Imaging Evaluation (DETERMINE) Trial”. In: *J. Cardiovasc. Electrophysiol.* 20.9 (Sept. 2009), pp. 982–987. DOI: 10.1111/j.1540-8167.2009.01503.x.

[63] T. Kam Ho. “The random subspace method for constructing decision forests”. In: *IEEE Trans. Pattern Anal. Mach. Intell.* 20.8 (1998), pp. 832–844. DOI: 10.1109/34.709601.

[64] J. Kittler, M. Hatef, R. Duin, and J. Matas. “On combining classifiers”. In: *IEEE Trans. Pattern Anal. Mach. Intell.* 20.3 (Mar. 1998), pp. 226–239. DOI: 10.1109/34.667881.

[65] L. I. Kuncheva and C. J. Whitaker. “Measures of Diversity in Classifier Ensembles and Their Relationship with the Ensemble Accuracy”. In: *Mach. Learn.* 51.2 (2003), pp. 181–207. DOI: 10.1023/A:1022859003006.

[66] O.-W. Kwon and T.-W. Lee. “Phoneme recognition using ICA-based feature extraction and transformation”. In: *Signal Processing* 84.6 (June 2004), pp. 1005–1019. DOI: 10.1016/j.sigpro.2004.03.004.

[67] L. Lam. "Classifier Combinations: Implementations and Theoretical Issues". In: *Mult. Classif. Syst.* 2000, pp. 77–86. DOI: 10.1007/3-540-45014-9{_}7.

[68] D. J. Lee et al. "Determination of carotid plaque risk by ultrasonic tissue characterization". In: *Ultrasound Med. Biol.* 24.9 (Dec. 1998), pp. 1291–1299. DOI: 10.1016/S0301-5629(98)00125-2.

[69] K.-H. Liu, B. Li, J. Zhang, and J.-X. Du. "Ensemble component selection for improving ICA based microarray data prediction models". In: *Pattern Recognit.* 42.7 (July 2009), pp. 1274–1283. DOI: 10.1016/j.patcog.2009.01.021.

[70] F. L. Lizzi, M. Greenebaum, E. J. Feleppa, M. Elbaum, and D. J. Coleman. "Theoretical framework for spectrum analysis in ultrasonic tissue characterization." In: *J. Acoust. Soc. Am.* 73.4 (1983), pp. 1366–73.

[71] F. Lizzi, M. Ostromogilsky, E. Feleppa, M. Rorke, and M. Yaremko. "Relationship of Ultrasonic Spectral Parameters to Features of Tissue Microstructure". In: *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 34.3 (May 1987), pp. 319–329. DOI: 10.1109/T-UFFC.1987.26950.

[72] F. L. Lizzi, E. J. Feleppa, S. Kaisar Alam, and C. X. Deng. "Ultrasonic spectrum analysis for tissue evaluation". In: *Pattern Recognit. Lett.* 24.4-5 (Feb. 2003), pp. 637–658. DOI: 10.1016/S0167-8655(02)00172-1.

[73] R. O. Mada, J. Duchenne, and J.-U. Voigt. "Tissue Doppler, Strain and Strain Rate in ischemic heart disease, How I do it". In: *Cardiovasc. Ultrasound* 12.1 (2014), p. 38. DOI: 10.1186/1476-7120-12-38.

[74] S. Maggio, A. Palladini, L. De Marchi, M. Alessandrini, N. Speciale, and G. Masetti. "Predictive Deconvolution and Hybrid Feature Selection for Computer-Aided Detection of Prostate Cancer". In: *IEEE Trans. Med. Imaging* 29.2 (Feb. 2010), pp. 455–464. DOI: 10.1109/TMI.2009.2034517.

[75] K. Mardia. “Statistics of Directional Data”. In: *J. R. Stat. Soc. Ser. B-Methodological* 37.3 (1975), pp. 349–393. DOI: 10.2307/2984782.

[76] L. Marks, S. Young, and S. Natarajan. “MRI-ultrasound fusion for guidance of targeted prostate biopsy.” In: *Curr. Opin. Urol.* 23.1 (2013), pp. 43–50. DOI: 10.1097/MOU.0b013e32835ad3ee.

[77] L. Masotti, E. Biagi, S. Granchi, L. Breschi, E. Magrini, and F. Di Lorenzo. “Clinical test of rules (rules: radiofrequency ultrasonic local estimators)”. In: *IEEE Ultrason. Symp. 2004.* Vol. 3. IEEE, pp. 2173–2176. DOI: 10.1109/ULTSYM.2004.1418269.

[78] L. Masotti, E. Biagi, S. Granchi, L. Breschi, E. Magrini, and F. Di Lorenzo. “Tissue differentiation based on radiofrequency echographic signal local spectral content”. In: *IEEE Symp. Ultrason. 2003.* IEEE, pp. 1030–1033. DOI: 10.1109/ULTSYM.2003.1293575.

[79] E. M. McMahon, J. Korinek, S. Yoshifuku, P. P. Sengupta, A. Manduca, and M. Belohlavek. “Classification of acute myocardial ischemia by artificial neural network using echocardiographic strain waveforms”. In: *Comput. Biol. Med.* 38.4 (Apr. 2008), pp. 416–424. DOI: 10.1016/j.compbiomed.2008.01.003.

[80] J. E. McNeal. “The zonal anatomy of the prostate.” In: *Prostate* 2.1 (1981), pp. 35–49. DOI: 10.1055/s-2008-1075885.

[81] J. E. McNeal, E. a. Redwine, F. S. Freiha, and T. a. Stamey. *Zonal Distribution of Prostatic Adenocarcinoma.* 1988. DOI: 10.1097/00000478-198812000-00001.

[82] Y. Mitani and Y. Hamamoto. “A local mean-based nonparametric classifier”. In: *Pattern Recognit. Lett.* 27.10 (July 2006), pp. 1151–1159. DOI: 10.1016/j.patrec.2005.12.016.

[83] G. F. Mitchell, G. a. Lamas, D. E. Vaughan, and M. a. Pfeffer. “Left ventricular remodeling in the year after first anterior myocardial infarction: A quantitative analysis of contractile

segment lengths and ventricular shape". In: *J. Am. Coll. Cardiol.* 19.6 (May 1992), pp. 1136–1144. DOI: 10.1016/0735-1097(92)90314-D.

[84] S. Mohamed and M. Salama. "Prostate Cancer Spectral Multifeature Analysis Using TRUS Images". In: *IEEE Trans. Med. Imaging* 27.4 (Apr. 2008), pp. 548–556. DOI: 10.1109/TMI.2007.911547.

[85] P. Mohana Shankar. "A general statistical model for ultrasonic backscattering from tissues". In: *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 47.3 (May 2000), pp. 727–736. DOI: 10.1109/58.842062.

[86] R. Molthen, P. Shankar, and J. Reid. "Characterization of ultrasonic B-scans using non-rayleigh statistics". In: *Ultrasound Med. Biol.* 21.2 (Jan. 1995), pp. 161–170. DOI: 10.1016/S0301-5629(94)00105-7.

[87] M. Moradi, P. Mousavi, A. Boag, E. Sauerbrei, D. Siemens, and P. Abolmaesumi. "Augmenting Detection of Prostate Cancer in Transrectal Ultrasound Images Using SVM and RF Time Series". In: *IEEE Trans. Biomed. Eng.* 56.9 (Sept. 2009), pp. 2214–2224. DOI: 10.1109/TBME.2008.2009766.

[88] A. Nair, B. D. Kuban, E. M. Tuzcu, P. Schoenhagen, S. E. Nissen, and D. G. Vince. "Coronary Plaque Classification With Intravascular Ultrasound Radiofrequency Data Analysis". In: *Circulation* 106.17 (Oct. 2002), pp. 2200–2206. DOI: 10.1161/01.CIR.0000035654.18341.5E.

[89] L. Parsons, E. Haque, and H. Liu. "Subspace clustering for high dimensional data". In: *ACM SIGKDD Explor. Newsl.* 6.1 (June 2004), pp. 90–105. DOI: 10.1145/1007730.1007731.

[90] G. Perk, P. a. Tunick, and I. Kronzon. "Non-Doppler Two-dimensional Strain Imaging by Echocardiography, From Technical Considerations to Clinical Applications". In: *J. Am. Soc. Echocardiogr.* 20.3 (Mar. 2007), pp. 234–243. DOI: 10.1016/j.echo.2006.08.023.

[91] D. Perperidis, R. Mohiaddin, and D. Rueckert. "Construction of a 4D Statistical Atlas of the Cardiac Anatomy and Its Use in Classification". In: *Lect. Notes Comput. Sci.* Vol. 3750 LNCS. 2005, pp. 402–410. DOI: 10.1007/11566489{_}50.

[92] *Physiology Notes*. URL: http://www.physiologynotes.com/stages-of-the-cardiac-cycle/.

[93] R. Polikar. "Ensemble based systems in decision making". In: *IEEE Circuits Syst. Mag.* 6.3 (2006), pp. 21–45. DOI: 10.1109/MCAS.2006.1688199.

[94] J. Raja, N. Ramachandran, G. Munneke, and U. Patel. "Current status of transrectal ultrasound-guided prostate biopsy in the diagnosis of prostate cancer". In: *Clin. Radiol.* 61.2 (Feb. 2006), pp. 142–153. DOI: 10.1016/j.crad.2005.10.002.

[95] F. J. Rohlf and D. Slice. "Extensions of the Procrustes Method for the Optimal Superimposition of Landmarks". In: *Syst. Zool.* 39.1 (Mar. 1990). DOI: 10.2307/2992207.

[96] D. Ruta and B. Gabrys. "Classifier selection for majority voting". In: *Inf. Fusion* 6.1 (Mar. 2005), pp. 63–81. DOI: 10.1016/j.inffus.2004.04.008.

[97] M. Scabia, E. Biagi, and L. Masotti. "Hardware and software platform for real-time processing and visualization of echographic radiofrequency signals". In: *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 49.10 (Oct. 2002), pp. 1444–1452. DOI: 10.1109/TUFFC.2002.1041086.

[98] U. Scheipers, H. Ermert, H.-J. Sommerfeld, M. Garcia-Schürmann, T. Senge, and S. Philippou. "Ultrasonic multifeature tissue characterization for prostate diagnostics". In: *Ultrasound Med. Biol.* 29.8 (Aug. 2003), pp. 1137–1149. DOI: 10.1016/S0301-5629(03)00062-0.

[99] G. Schmitz, H. Ermert, and T. Senge. "Tissue-characterization of the prostate using radio frequency ultrasonic signals". In: *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 46.1 (Jan. 1999), pp. 126–138. DOI: 10.1109/58.741523.

[100] B. Schölkopf, A. Smola, and K.-R. Müller. "Nonlinear Component Analysis as a Kernel Eigenvalue Problem". In: *Neural Comput.* 10.5 (July 1998), pp. 1299–1319. DOI: 10.1162/089976698300017467.

[101] P. M. Shankar. "A model for ultrasonic scattering from tissues based on the K distribution". In: *Phys. Med. Biol.* 40.10 (Oct. 1995), pp. 1633–1649. DOI: 10.1088/0031-9155/40/10/006.

[102] P. Shankar. "Ultrasonic tissue characterization using a generalized Nakagami model". In: *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 48.6 (2001), pp. 1716–1720. DOI: 10.1109/58.971725.

[103] P. Smets. "Decision making in the TBM: the necessity of the pignistic transformation". In: *Int. J. Approx. Reason.* 38.2 (Feb. 2005), pp. 133–147. DOI: 10.1016/j.ijar.2004.05.003.

[104] M. B. Stegmann, K. Sjöstrand, and R. Larsen. "Sparse modeling of landmark and texture variability using the orthomax criterion". In: *Soc. Photo-Optical Instrum. Eng. Conf. Ser.* Ed. by J. M. Reinhardt and J. P. W. Pluim. Vol. 6144. Mar. 2006, pp. 1–12. DOI: 10.1117/12.651293.

[105] A. Suinesiaputra et al. "Automated Detection of Regional Wall Motion Abnormalities Based on a Statistical Model Applied to Multislice Short-Axis Cardiac MR Images". In: *IEEE Trans. Med. Imaging* 28.4 (Apr. 2009), pp. 595–607. DOI: 10.1109/TMI.2008.2008966.

[106] G. R. Sutherland, G. Di Salvo, P. Claus, J. D'hooge, and B. Bijnens. "Strain and strain rate imaging: a new clinical approach to quantifying regional myocardial function". In: *J. Am. Soc. Echocardiogr.* 17.7 (July 2004), pp. 788–802. DOI: 10.1016/j.echo.2004.03.027.

[107] T. L. Szabo. *Diagnostic Ultrasound Imaging: Inside Out*. Elsevier, 2014. DOI: 10.1016/B978-0-12-396487-8.00005-7.

[108] M. Tabassian, F. Galluzzo, L. De Marchi, N. Speciale, G. Masetti, and N. Testoni. "Soft-label reinforced rtCAB for guided prostate tissue sampling". In: *2013 IEEE Int. Ultrason. Symp.* IEEE, July 2013, pp. 880–883. DOI: 10.1109/ULTSYM.2013.0226.

[109] M. Tabassian, R. Ghaderi, and R. Ebrahimpour. "Combining complementary information sources in the Dempster-Shafer framework for solving classification problems with imperfect labels". In: *Knowledge-Based Syst.* 27 (Mar. 2012), pp. 92–102. DOI: 10.1016/j.knosys.2011.10.010.

[110] M. Tabassian et al. "Principal Component Analysis for the Classification of Cardiac Motion Abnormalities Based on Echocardiographic Strain and Strain Rate Imaging". In: vol. 9126. Lecture Notes in Computer Science. 2015, pp. 83–90. DOI: 10.1007/978-3-319-20309-6_10.

[111] J. B. Tenenbaum, V. De Silva, and J. C. Langford. "A Global Geometric Framework for Nonlinear Dimensionality Reduction". In: *Science* 290.5500 (Dec. 2000), pp. 2319–2323. DOI: 10.1126/science.290.5500.2319.

[112] N. Testoni, S. Maggio, F. Galluzzo, L. De Marchi, and N. Speciale. "rtCAB: A tool for reducing unnecessary prostate biopsy cores". In: *2010 IEEE Int. Ultrason. Symp.* IEEE, Oct. 2010, pp. 49–52. DOI: 10.1109/ULTSYM.2010.5935788.

[113] C. Tong, A. Palazoglu, and X. Yan. "Improved ICA for process monitoring based on ensemble learning and Bayesian inference". In: *Chemom. Intell. Lab. Syst.* 135 (July 2014), pp. 141–149. DOI: 10.1016/j.chemolab.2014.04.012.

[114] P. Tortoli, L. Bassi, E. Boni, A. Dallai, F. Guidi, and S. Ricci. "ULA-OP: an advanced open platform for ultrasound research". In: *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 56.10 (Oct. 2009), pp. 2207–2216. DOI: 10.1109/TUFFC.2009.1303.

[115] T. Tuthill. "Deviations from Rayleigh statistics in ultrasonic speckle". In: *Ultrason. Imaging* 10.2 (Apr. 1988), pp. 81–89. DOI: 10.1016/0161-7346(88)90051-X.

[116] M. Unser. “Sum and Difference Histograms for Texture Classification”. In: *IEEE Trans. Pattern Anal. Mach. Intell.* PAMI-8.1 (Jan. 1986), pp. 118–125. DOI: 10.1109/TPAMI.1986.4767760.

[117] V. N. Vapnik. *Statistical Learning Theory*. John Wiley Sons, 1998. DOI: 10.2307/1271368.

[118] M. Varma and A. Zisserman. “A Statistical Approach to Material Classification Using Image Patch Exemplars”. In: *IEEE Trans. Pattern Anal. Mach. Intell.* 31.11 (Nov. 2009), pp. 2032–2047. DOI: 10.1109/TPAMI.2008.182.

[119] Z. Vered et al. “Quantitative ultrasonic tissue characterization with real-time integrated backscatter imaging in normal human subjects and in patients with dilated cardiomyopathy”. In: *Circulation* 76.5 (Nov. 1987), pp. 1067–1073. DOI: 10.1161/01.CIR.76.5.1067.

[120] R. F. Wagner, M. F. Insana, and D. G. Brown. “Statistical properties of radio-frequency and envelope-detected signals with applications to medical ultrasound”. In: *J. Opt. Soc. Am. A* 4.5 (May 1987), p. 910. DOI: 10.1364/JOSAA.4.000910.

[121] S. A. Wickline, E. D. Verdonk, A. K. Wong, R. K. Shepard, and J. G. Miller. “Structural remodeling of human myocardial tissue after infarction. Quantification with ultrasonic backscatter.” In: *Circulation* 85.1 (1992), pp. 259–68. DOI: 10.1161/01.CIR.85.1.259.

[122] T. Windeatt and R. Ghaderi. “Coding and decoding strategies for multi-class learning problems”. In: *Inf. Fusion* 4.1 (Mar. 2003), pp. 11–21. DOI: 10.1016/S1566-2535(02)00101-X.

[123] D. H. Ye, B. Desjardins, J. Hamm, H. Litt, and K. M. Pohl. “Regional Manifold Learning for Disease Classification”. In: *IEEE Trans. Med. Imaging* 33.6 (June 2014), pp. 1236–1247. DOI: 10.1109/TMI.2014.2305751.

[124] X. Zhang et al. “Atlas-Based Quantification of Cardiac Remodeling Due to Myocardial Infarction”. In: *PLoS One* 9.10 (Oct. 2014). Ed. by W. R. Bauer, e110243. DOI: 10.1371/journal.pone.0110243.

List of Publications

[1] M. Tabassian, F. Galluzzo, L. De Marchi, N. Speciale, G. Masetti, and N. Testoni, "Soft-Label Reinforced rtCAB for Guided Prostate Tissue Sampling", *IEEE Int. Ultrason. Symp. 2013*, pp. 880–883.

[2] M. Tabassian, N. Testoni, L. De Marchi, F. Galluzzo, N. Speciale and G. Masetti, "Learning Features from Medical Radiofrequency Ultrasonic Signals by Independent Component Analysis", *IEEE 27th Int. Symp. Comput. Med. Syst. 2014*, pp. 62–65.

[3] M. Tabassian, M. Alessandrini, L. De Marchi, G. Masetti, N. Cauwenberghs, T. Kouznetsova, and J. D'hooge, "Principal Component Analysis for the Classification of Cardiac Motion Abnormalities Based on Echocardiographic Strain and Strain Rate Imaging", *The 8th Int. Conf. on Functional Imaging and Modeling of the Heart (FIMH), 2015*, pp. 83–90.

[4] M. Tabassian, M. Alessandrini, L. Herbots, O. Mirea, J. Engvall, L. De Marchi, G. Masetti, and J. D'hooge, "Automatic Detection of Ischemic Myocardium by Spatio-Temporal Analysis of Echocardiographic Strain and Strain Rate Curves", *IEEE Int. Ultrason. Symp. 2015*.

[5] M. Tabassian, M. Alessandrini, P. Claes, L. De Marchi, D. Vandermeulen, G. Masetti and J. D'hooge, "Automatic Detection of Myocardial Infarction Through a Global Shape Feature based on Local

Statistical Modeling", *Statistical Atlases and Computational Modeling of the Heart (STACOM) workshop, MICCAI 2015.*

[6] M. Tabassian, L. De Marchi, N. Testoni, F. Galluzzo, J. D'hooge and G. Masetti, "A Novel Ensemble Scheme Based on ICA Dictionaries for the Classification of Medical Radiofrequency Ultrasonic Signals", *Under preparation.*

[7] M. Tabassian, M. Alessandrini, L. Herbots, O. Mirea, J. Engvall, L. De Marchi, G. Masetti, and J. D'hooge, "Spatio-Temporal Analysis of Echocardiographic Deformation Curves for Automatic Detection of Myocardial Infarction", *Under preparation.*